THE ROLE OF AGGLOMERATION AND DISPERSION FORCES ON THE SPATIAL DISTRIBUTION OF ECONOMIC ACTIVITY IN CHILE: AN EMPIRICAL ANALYSIS OF SPATIAL WAGE DISTRIBUTION AND REGIONAL HOUSING PRICES

BY

DUSAN JAN CARLOS PAREDES ARAYA

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Agricultural and Applied Economics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2012

Urbana, Illinois

Doctoral Committee:

Professor Geoffrey J.D. Hewings, Chair
Professor Kathy Baylis
Professor Anil K. Bera
Professor Edward Feser
Professor Alex Winter-Nelson

Abstract

This thesis begins with an **Introduction** oriented to guide the reader in two main aspects. First, Chile presents an excessive level of spatial concentration around the Metropolitan Region of Santiago (MR). Second, this spatial concentration is critical for the case of wages. In spite of the relevance, the **Introduction** shows the lack of discussion to understand the forces affecting this spatial distribution. In this sense, the **Introduction** proposes "[This thesis]covers this gap offering three essays oriented to understand what factors are generating this wage distribution and the consequences of the excessive spatial concentration on the MR, such as the congestion cost associated with the housing prices". The **Chapter 2** is the first approach to estimate the role of the second nature on the regional wages through the estimation of two NEG models. This chapter suggests the important role of the second nature, but the evidence is weak. The lack of evidence provided by **Chapter 2** is attributed to the fact that NEG captures one of the potential set of sources behind the wage differentials. This problem motivates the development of the **Chapter 3**. This chapter proposes different causal mechanisms that modify the theoretical NEG model to represent the additional sources. A new empirical approach is provided to evaluate the role of alternative theories such as the role of amenities. Finally, the **Chapter 4** shows a different side of the spatial concentration which is related with the dispersion forces. This section proposes a novel methodology to estimate a housing price index at regional level. This new methodology allows to compare the price of heterogeneous goods across the space, being a contribution for regional analysis of dispersion forces. While the regions with initial endowments, such as copper and salmon, show high level of housing prices; MR does not present the highest level of housing price. This result suggests that the case of Chile seems to be managed by agglomeration forces stronger than dispersion forces. Moreover, the concentration of high wages around the MR could be higher, increasing the magnitude of spatial inequality for the case of Chile.

Para mi madre por darme lo que incluso ella no posea.

Acknowledgments

This thesis is the output of different people who supported me during this long process. First, I would like to my advisor who always push me to move forward, Patricio Aroca, who is not only an incredible advisor, he is also my friend. Secondly, I thank to my advisor Andrew Isserman who regretably passed away during my thesis process, but with his wonderful advices I already had the tools to finish this stage of my life. Finally, I wish to say thanks to Geoffrey Hewings, a person who support me during this last stage and who always believed in my work. Thank you for all of you.

Table of Contents

Chapter 1

Introduction

The spatial distribution of wages does not follow a random process: high wages are generally concentrated in a core, while the low wages are relegated to peripheral zones. This stylized fact has considerable policy implications for developing countries who have to deal with higher degrees of urban primacy than developed countries and where the excessive spatial concentration is linked with negative consequences such as low productivity, social dissatisfaction, negative externalities and high levels of regional inflation manifested in housing prices (Puga, 2002, Henderson, 2003). An example of this reality is Chile which has one of the highest levels of income inequality in the world (Ferreira and Litchfield, 1999, Contreras, 2003) and whose inequality can be easily extended toward a spatial dimension. Aroca and Bosch (2000), used spatial econometric tools to test the hypothesis of convergence in Gross Regional Product (GRP) per capita between 1990-98; their results suggested that poorer regions are not catching the richer regions[1]. This spatial divergence hypothesis is reinforced by Echeverria and Gopinath (2007) who noted that the Metropolitan Region of Santiago (MR) counts with 2% of the total territory, 50% of the total production and 40% of the total population[2]. However, a crucial point seems to be the persistence in the spatial inequality of wages during the last decade. According to Aroca (2009), in 1992, only two regions had a GRP higher than the MR, but the MR accounted with the highest level of income per capita. By 2003, all the regions showed a positive growth of GRP relative to the MR, but none of them had a wage rate higher than the MR, indicating an "absortion" of wages from the MR. Notwithstanding of the empirical evidence that has revealed the existence of a spatial distribution of wages in Chile, there is limited understanding about the factors shaping this spatial concentration. This thesis covers this gap offering three essays oriented to undertand what factors are generating this wage distribution and the consequences of the excessive spatial concentration on the regions, such as the congestion cost associated with the housing prices.

[1]As the reader can see the literature uses the word "region" to make reference to any spatial level lower than country. In Chile, the word "region" coincides with the political name of the second-level spatial unit. To be consistent with the literature, I use the word region to indicate the administrative unit of Chile.

[2]Before October, 2007, Chile was divided into 13 regions including the MR. In October, 2001, two new regions were added, making 15 regions including the MR. Regions are divided into 54 provinces and provinces are further divided into 346 communes. Communes can be considered equivalent to counties for United States.

Combes et al. (2008a) identify three sources to explain the spatial disparities in wages, where all of them affect the wage through the productivity of the workers. The first one is represented by the skills associated with each worker. This source can be appreciated through the role played by industries that require a specialized workforce, attracting workers with high skills and pushing up the average wage with important consequences on the spatial inequality. The second source, namely initial endowments, affects the productivity through favorable conditions, such as the climate or natural resources. This source has been labeled as "first nature" and its role is relevant for developing countries where the natural resources play an important role in the economic structure. The third source comes from the interaction in a particular geographical space between workers and firms. This source was earlier described by Marshall (1890), but formalized by some branches inside of regional science such as Urban Economics (UE) or the New Economic Geography (NEG). This source has received several names, but the "second nature" (Krugman, 1993) and economics of agglomeration (Fujita and Thisse, 2002) are among the most accepted ones.

The literature has not identified explicitly which one of these sources affects the spatial distribution of wages in Chile, but some hints can be identified. For example Aroca (2009) suggests that the MR describes a particular dynamic relationship of GRP and wages during the last decade. The MR does not belong to the group of regions with high growth rates of GRP in Chile. However, the MR has increased its wage growth more than any region in Chile. To explain this asymmetric relationship between GRP and wage, this thesis formulates a hypothesis based on the role played by the second nature. The MR is not characterized by the exploitation of natural resources, ports or good conditions of climate (first nature), but the MR accounts for the highest level of spatial concentration among the Chilean regions, suggesting a wage concentration based on the interaction of economic agents. This stylized fact motivates this thesis to evaluate if , controlling by first nature and the skill of the labor force, the second nature can explain the current spatial distribution of wages. If we cannot reject this hypothesis, then the continuing spatial concentration would imply the negative consequences described previously.

In order to evaluate the role played by the second nature, economic theory provides several options theoretical backgrounds: i) Urban Economics (UE) (Alonso, 1964) and ii) New Economic Geography (NEG) (Krugman, 1991). While the UE is suitable to analyze urban structures within cities, the NEG analyzes the location decision at the interregional level[3]. An alternative differentiation could be based on "how" both theories can explain the agglomeration process. According to Krugman (1991), the UE explain the agglomeration process with the incorporation of localized production externalities; however there is not a

[3]Thisse (2010) critiques this separation as an arbitrary one arguing that interregional systems are supported by urban structures and therefore a differentiation based on spatial scale is diffuse. I prefer this taxonomy because it has been adopted by the most of the regional science's literature.

clear microeconomic understanding about how they are generated. On the other hand, Fujita and Thisse (2009) claim that the NEG is the first general equilibrium model that explain "how" spatial inequalities emerge from the interaction among increasing returns, transport cost and imperfect competition. This paper is motivated by the second definition for two reasons. First, the scarce information that is available at the regional or commune level, but it is not available at the city-level making intra urban analysis difficult. Moreover, the differentials in the wage distribution exist at the regional level, suggesting that this may be the most appropriate scale to explore the NEG theoretical framework.

Putting in order the concepts described previously, this thesis evaluates whether or not the second nature can explain the spatial distribution of wages in Chile using the causal mechanism proposed by the NEG. To carry out this objective, this thesis presents three empirical essays applied to the case of Chile. The first essay analyzes whether or not the spatial distribution of wages at the communal level can be modeled through a empirical estimation of the NEG's extensions proposed by Helpman (1998) and Sudekum (2006) (hereafter referred to as the Helpman and Sudekum's models). Particularly, this theoretical background proposes that the spatial distribution of wages may be considered as the consequence of a tension between the market access-effect (agglomeration force) and housing-price effect (dispersion force), both examples of the second nature. While Helpman's model has been already estimated (Brakman et al., 2004, Kiso, 2005, Mion, 2004, Hanson, 2005), this paper is the first one estimating Sudekum's model. Estimating both models, this paper provides evidence to locate the Chilean economy in the "bell shape". This exercise is useful to evaluate what level of spatial concentration we should expect in the future[4] The results suggest that the spatial distribution of wages in Chile is driven by stronger market-access than housing-price effects. Moreover, the Chilean case seems to fit with a developing country (left side of the bell shape). However, this hypothesis is very weak with respect to the empirical analysis for other countries. The main hypothesis for this weak result is sustained by the strong role played by first nature, which is not a dimension considered by the NEG.

The evidence provided by the first essay to argue for the existence of a second nature effect is weaken (probably) by the strong role played by the first nature. However, a second explanation is related with the multiple set of alternatives sources, alternative to the NEG, to explain the spatial distribution of wages. The second chapter proposes several considerations to improve the weakness identified on the previous chapter. First, the estimation at an aggregated level, namely communal level, does not help to control for the additional sources of wage variation such as the first nature. Given the access to micro data for the case of Chile, the second essay will improve the methodological approach, estimating the model with multilevel

[4]A deep explanation about this point is commented on the Chapter 1.

analysis. Second, the first essay is focused on a particular year, 2006, but a new data set is available for this new articule. The second essay improves this weakness by using data for the year 2009 to evaluate the role played by the second nature and natural resources in the spatial distribution of wages. Finally, I carry out my identification strategy following two steps. In the first step I define what the potential sources behind the spatial wage differential and, secondly, I propose a modified economic model to capture these sources. In order to capture both dimensions, this chapter propose a competition between New Economic Geography and Spatial Equilibrium model to evaluate what causal mechanism show a higher fit for the Chilean case.

The first and second essays evaluate the role played by the second nature in the composition of the spatial distribution of wages, but they do not provide evidence to evaluate the consequences of a particular distribution on the rest of economic dimensions. To carry out this evaluation, this thesis evaluates the congestion costs associated with high levels of agglomeration, namely those associated with housing price. This goal presents additional challenges. From an axiomatic perspective, a price index must consider the relative cost of accessing a similar housing basket in different locations. This assumption is extremely complex because the heterogeneity of houses as well as the geographical requirements make it difficult to consider of a similar houses. Using matching estimator methods, a regional housing price is built for each one of the regions. Regions with high wages are positively correlated with high housing prices, but housing prices are higher for those agglomerations generated by the first nature.

These three essays contribute to an evaluation of the sources of the spatial concentration of wages and what are their potential consequences on this spatial distribution. Assuming that the excessive spatial concentration of wages has several negative consequences on equity and efficiency, then it is very important to identify with a causal mechanism (NEG) those forces. Particularly, Chile shows a high and persistent spatial concentration of economic activity and population around the MR. Simultaneously, this spatial concentration can promote more productivity (second nature) and higher wages around the MR, creating a pattern core-periphery in Chile. From a policy perspective, the agglomeration of wages around the second nature is extremely complex to evaluate. Moreover, this research puts in evidence that some economic policies must be analyzed considering the complete spatial inter-relations among regions. Otherwise, the excessive spatial concentration of Chile can affect seriously the country's economic development and future economic growth.

Chapter 2

Can NEG explain the spatial distribution of wages in a developing country: Evidence from Chile

Road map for Chapter 2: The Section 2.2 indicates the alarming case of spatial concentration of wages for Chile. This section discusses the particular role played by the combination of the first nature and second nature, a typical case of a developing country. Section 2.3 shows the causal mechanism to explain the case of Chile, namely the New Economic Geography. The market access and housing price are highlighted as crucial factors to evaluate the location of the economic activity. The section formally details the two estimated models : Sudekum and Helpman's model. The Section 2.5 proposes and argue the choice of the variables used to estimate the wage equation. Simultaneously, a detailed discussion is carried out to evaluate the proper function form of the transportation cost. Finally, Sections 2.6 and 2.7 highlight the role played by the second nature to explain the case of Chile; however some disadvantages about the model are discussed to motivate the Chapter 3 of this thesis.

2.1 Introduction

Developing countries are characterized by alarming levels of spatial concentration (Puga, 1998, Rosenthal and Strange, 2004). Chile illustrates this case; several descriptive studies show high spatial concentration of wages around the Metropolitan Region (MR), which is reinforced by tremendous regional disparities in several economic variables, such as GDP and population[1] (Aroca, 2009). This situation could become a complex problem considering that the literature identifies negative consequences of excessive spatial concentration, such as low productivity, social dissatisfaction, negative externalities and high levels of regional inflation, among others (Henderson, 2003, Armstrong and Taylor, 2000). In spite of this, there is no literature dedicated to formally evaluating whether Chile is characterized by excessive spatial concentration of wages. This paper provides a first estimation of an economic theory to evaluate whether Chile presents this situation. The theoretical background to evaluate the case of Chile is provided by the New Economic Geography (NEG), an economic theory designed to explain the spatial concentration of wages. This evaluation is focused on two

[1] A deeper analysis of the Chilean case is detailed in the next section.

critical questions: 1) whether the NEG explains the spatial distribution of wages in a developing country like Chile, and 2) how the NEG can be used to infer information about future levels of spatial concentration of wages in Chile. Through these questions, this paper provides an evaluation necessary for the design of future regional policies oriented to reducing the potential negative consequences of excessive spatial concentration.

To make the existence of the spatial distribution of wages evident, we must first recognize the potential sources of this concentration. Combes et al. (2008a) suggested three sources to explain spatial wage disparities: the skill composition of the workforce, non-human endowments (first nature) and local interaction (second nature[2]). While the NEG explains the disparities produced by second nature, this theory does not make reference to the other two sources. Of the sources left out, first nature is a key source to explain the spatial distribution of wages in developing countries. Such countries are mainly exporters of natural resources (a type of non-human endowment), which presents a paradox: on the one hand, developing countries show alarming levels of spatial concentration, the explanation of which is the main objective of the NEG. On the other hand, developing countries presents wage differentials linked to the existence of natural resources, a source left out by the NEG. This paradox raises the question of whether the NEG can replicate the spatial distribution of wages in countries where several sources of wage disparities are mixed together. This paper answers this question by estimating the NEG for the case of Chile, a country with one of the highest levels of income inequality in the world, and where a fundamental role is played by first nature.

Given the lack of empirical discussion about how the NEG explains the spatial disparities of wages in developing countries, there is less information on how these spatial disparities evolve over time. This paper proposes a simple method to predict future levels of spatial concentration using the well known “bell-shaped curve of spatial development” (Puga, 1999). Bell-shaped spatial development is characterized by three stages: dispersion, concentration and dispersion again, as Figure (2.1) describes. In the early stage of development in a particular country, high transport costs encourage dispersed centers with similar spatial concentrations of economic activity. The reduction in transport costs encourages concentration and increases wage disparities, which could represent the current realities of many developing countries. At some point, negative externalities outweigh the benefits of concentration, leading to dispersion again. This last phenomenon could represent the situation of countries that have already considerably reduced transportation costs, such as the case for most developed countries.

The literature identifies a statistical regularity where developing countries and developed countries can be located on the left and right side, respectively, of a bell-shaped curve representing spatial development (Combes et al., 2008b). I used this intuition to develop a simple method to determine on what side Chile

[2]An early conceptualization of second nature was discussed by Marshall (1890), while the modern term “economics of agglomeration” was discussed by (Fujita and Thisse, 2002)

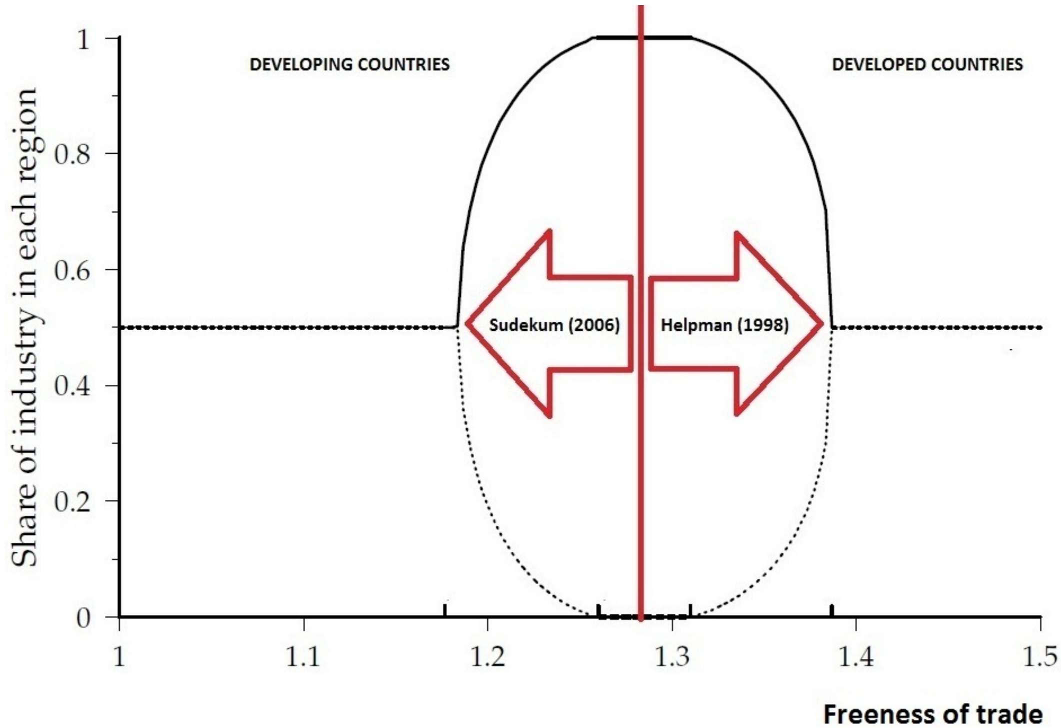

Figure 2.1: Bell-shape curve of spatial development.

is located. To carry out this objective, I applied two models, one developed by Helpman (1998) and the other by Sudekum (2006) (hereafter known respectively as Helpman's and Sudekum's models) to evaluate whether the spatial distribution of wages can be explained by the factors described by the NEG, namely the tension between the market access-effect (agglomeration force) and the housing-price effect (dispersion force). However, the two models can be assumed as two sides of bell-shaped spatial development because they represent different concentration processes, as described by Figure (2.1). Sudekum's model starts with total dispersion, and a high level of spatial agglomeration appears as an efficient equilibrium. Helpman's model starts with a high level of concentration and a dispersion phenomenon then takes place, replicating the case of developed countries such as the United States. I show that Helpman's model is a special case of Sudekum's model, and that the two can be seen to represent two sides of the bell-shaped curve. As far as I know, this paper is the first attempt to provide this kind of evaluation, which is crucial for developing countries.

Additionally, this paper presents several contributions to the standard empirical literature of the NEG. I contribute to reducing the empirical gap for developing countries by assessing an NEG model to analyze the spatial distribution of wages in Chile for 334 spatial units, called communes, representing the first estimation of the NEG in this country and at this spatial level. While Helpman's model has already been estimated (Brakman et al., 2004, Kiso, 2005, Mion, 2004, Hanson, 2005), this paper is the first to estimate Sudekum's model. This estimation provides two nested models to evaluate whether there will be higher levels of concentration in Chile in the future. This information, develop for small low spatial units is relevant for the development of regional policy to reduce potential disparities, especially in cases where there is limited information at more disaggregated spatial scales. While some data sets are available for developed countries (from a temporal perspective), this is not the case for developing countries. This paper proposes a novel way to carry out this evaluation, but care must be given in evaluating the outcomes because the lack of an explicit temporal dimension.

2.2 Spatial distribution of wages in Chile.

Chile has one of the highest levels of income inequality in the world (Ferreira and Litchfield, 1999, Contreras, 2003), and this inequality has a spatial dimension. Aroca and Bosch (2000) used spatial econometrics to reject the hypothesis of convergence in Gross Regional Product (GRP) per capita between 1990-98. This result was supported by Echeverria and Gopinath (2007) who noted that the Metropolitan Region of Santiago (MR), with only 2% of the total territory, accounts for 50% of production and 40% of the population[3]. According to Aroca (2009), by 1992 only two regions had a higher GRP than that of the MR, but the MR had the highest wage levels. By 2003, all the regions showed positive growth of GRP relative to the MR, but none of them had higher wages than those of the MR, indicating an "absorption" of wages by the MR.

This evidence indicates the existence of spatial inequality in per capita GRP and wages. Unlike other articles, this paper focuses on wages instead of GRP. From a statistical perspective, wage data are more readily available than GRP data for more spatial units. While Morande et al. (1997), Aroca and Bosch (2000) and Duncan and Fuentes (2006) estimated their models with GRP data available for 13 Chilean regions (regional level), I increased the size of the sample by using wage data for 334 communes (communal level). Moreover, the NEG assumes that wage differentials are generated by economies of agglomeration, or second nature, that are mostly localized at a low spatial scale (Rosenthal and Strange, 2004). Thus, the communal, rather than the regional level, properly represents these theoretical requirements. Additionally, Aroca (2009) identifies a spatial divergence in wages, with the wage of the MR increasing more rapidly than those in any other region. From this perspective, regional divergence in wages is a potential concern for policy makers, therefore empirical studies are needed to evaluate future policies to reduce spatial inequality. Finally the GRP may be earned but not necessarily spent or remain in a region. While the wage can be

[3]Before October, 2007, Chile was divided into 13 regions, including the MR. In October, 2001, two new regions were added, making 15 regions, including the MR. The regions are divided into 54 provinces and provinces are further divided into 346 communes. The communes can be considered equivalent to counties in the United States.

directly related with an expenditure on the local region[4].

The arguments described previously are supported by the spatial distribution of wages for 2006[5] (See Figure 2.2). Wages around the MR are between 95% and 132% of the mean national wage, with two small clusters of high wages located in the extreme north and south. The first is a mining cluster based on cooper mining, while the second in the south is termed the salmon cluster. Aroca and Bosch (2000) provided statistical evidence for the mining cluster, but they described the salmon cluster as being in an initial stage. Some years later, the salmon cluster has been established, making the spatial distribution of wages in Chile more uneven. However, the two clusters are not strong enough to reduce the core-periphery structure generated by the wages around the MR (see the zoom map of the MR in Figure 2.2).

[4]The literature has identified some wages-flows ouside regions; however this outflow is lower that the case of GRP.

[5]This spatial distribution was constructed using the Chilean Household Survey CASEN 2006. A complete description of the data is in Section 2.5.

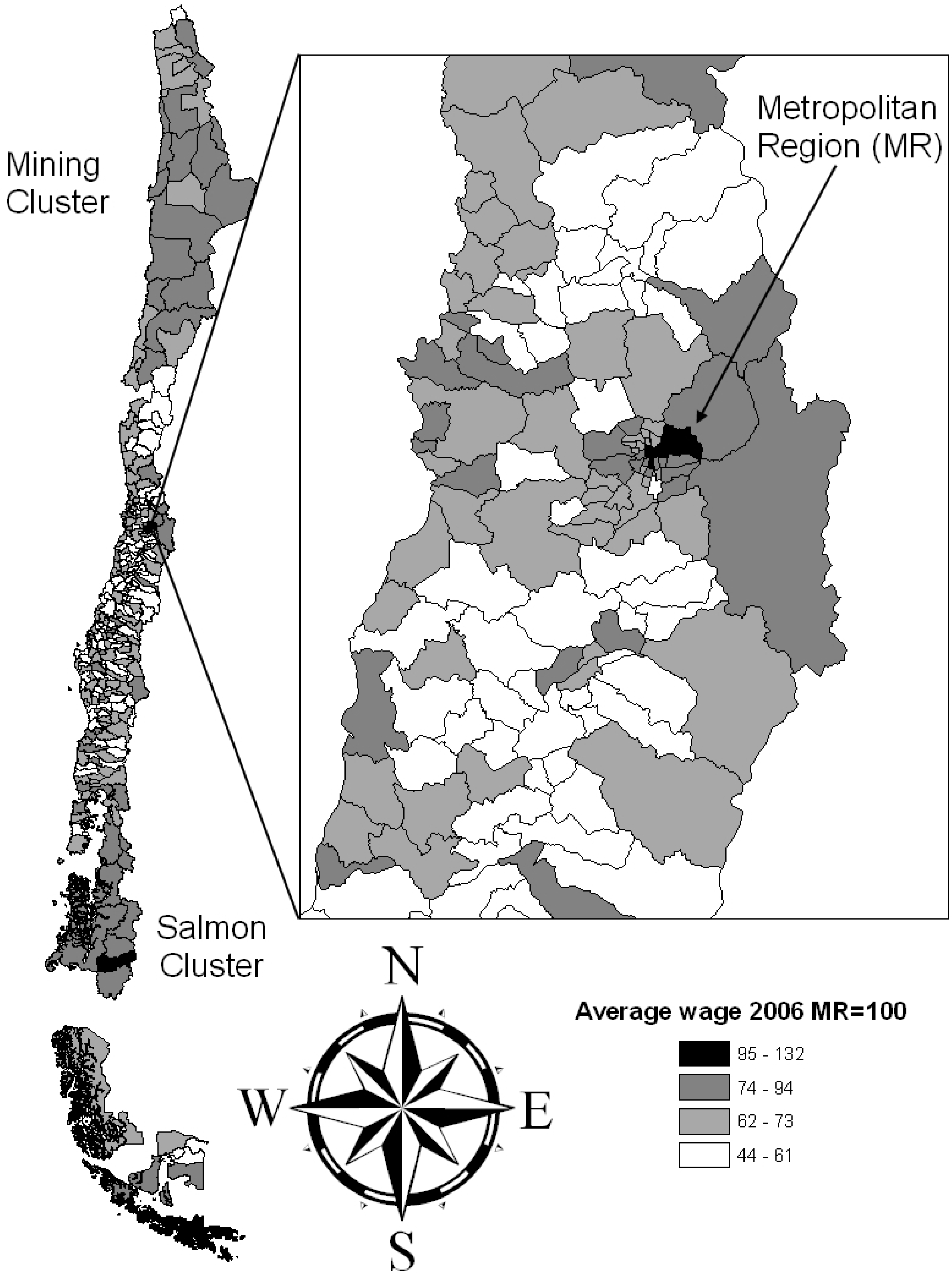

Figure 2.2: Spatial distribution of wages for 334 communes and zoom map for MR.

2.3 Causal relationship of the NEG for two regions.

The NEG is a highly nonlinear general equilibrium model in which the causal mechanism is not made clear solely by observing the equations. To avoid this problem, I follow Baldwin et al. (2003) who provided a simple example to illustrate the standard forces of the NEG: the market-access effect, the cost of living effect and the market-crowding effect, but I complement these with a new force: the housing-price effect. Let two symmetric regions be called the northern (N) and southern (S) regions, assuming a small migration of workers from S to N, which increases the relative market size of N[6]. Because of transport costs and increasing returns, firms prefer to locate near large markets following the initial migration of workers from S to N[7]. Migration of workers and firms represents the first agglomeration force, called the market-access effect or demand-linked circular causality, which increases the relative wage in N. The word circular is added because after the "first round" of migration, new jobs and higher wages are generated in N, attracting new workers again[8].

Growth in the number of firms in N increases the variety of industrial goods produced locally. All other things being equal, this increase reduces the industrial price index and consequently reduces the cost of living (COL)[9]. Reduction in the COL, ceteris paribus, increases relative real wages in N, attracting workers to the agglomeration. This process is called the cost-of-living effect or cost-linked circular causality. Again, the word circular is added because the migration of workers increases the market, attracting new firms and further reducing the COL. In summary, both agglomeration forces increase wages in N compared to S. However, the NEG also establishes forces working in the opposite direction.

The migration of firms and workers increases the level of competition in N, thus reducing profits. To compensate for the loss, firms must pay lower wages, reducing the relative wage in N. This dispersion force is called the market-crowding effect or the market-crowding dispersion force, which does not involve a circular process. The second dispersion force is called the housing-price effect, which was formalized by Helpman (1998) and Sudekum (2006). After the initial migration, N has a higher demand for housing, pushing up housing prices and reducing the incentives to stay agglomerated in N because real wages are lower than in S[10]. Whether one force is stronger than the other depends on transport costs.

[6] If the initial "random" migration seems to be improbable, then assume two asymmetrical regions. The causal mechanism works in a similar way.

[7] Increasing returns imply that a firm can reduce the average cost by producing large quantities. Thus, firms prefer to concentrate production in just one plant. In this scenario, firms are indifferent to the specific location (N or S). However, transportation costs imply that firms must pay to export industrial goods. Thus, firms prefer to be located in large markets to reduce transport costs.

[8] This circular process refers to the cumulative causation addressed by Myrdal (1957)

[9] The market is assumed to be in imperfect competition (Dixit and Stiglitz, 1977) and the competition is different from that of markets with perfect competition. Firms compete with varieties, therefore an increase in the number of firms implies a reduction in the industrial price index.

[10] A third dispersion force discussed by Fujita et al. (1999) is called the immobile-sector dispersion force. However, Baldwin et al. (2003) do not include it as a dispersion force. There are workers in the southern region who cannot migrate to the north.

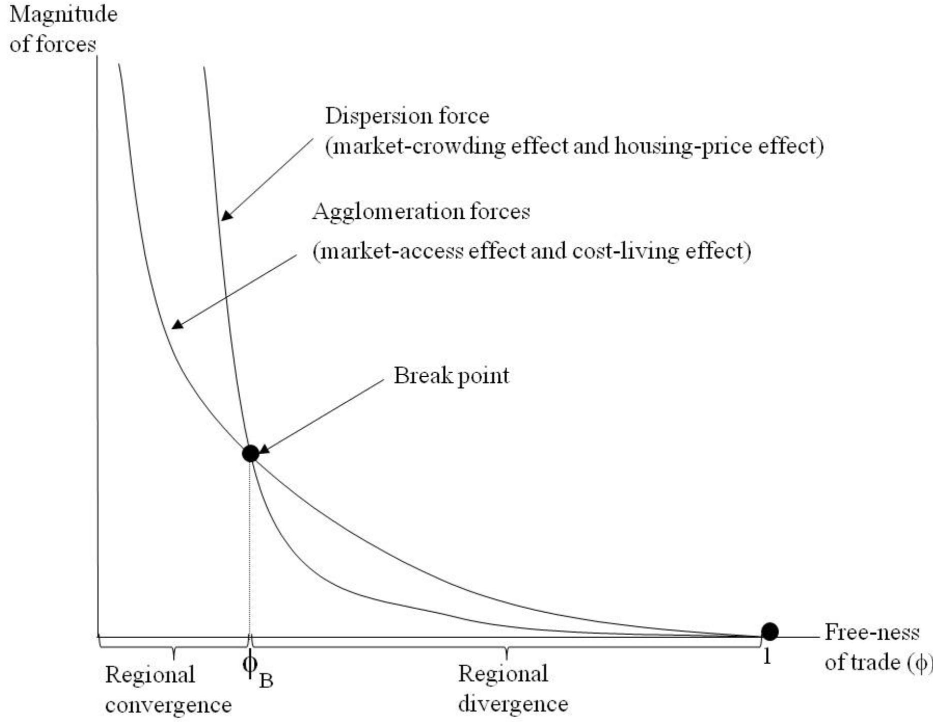

Figure 2.3: Agglomeration and dispersion forces erode with trade freeness (Baldwin et al., 2003).

A summary is presented in Figure 2.3, where the magnitude of the forces are a function of the freeness of trade ϕ. Transport costs have an inverse relationship to ϕ such that high transport costs imply a low level of ϕ (equal to 0) and low transport costs imply a high level of ϕ (equal to 1). An increase in ϕ reduces the magnitude of both forces, but dispersion forces decrease more rapidly than agglomeration forces. To analyze this difference, a mathematical analysis is needeed, but some intuitions can be derived from the previous example where agglomeration forces are characterized by a reinforcing process, while the dispersion forces are not. This implies that dispersion forces fall more rapidly than agglomeration forces with an increase in ϕ.

In Figure 2.3, the threshold ϕ_B is termed the "break-even point". To the left of ϕ_B (lower levels of ϕ), dispersion forces are stronger than agglomeration forces, shaping similar wages among spatial units. To the right of ϕ_B (higher levels of freeness of trade), agglomeration forces are stronger than dispersion forces, generating a spatial distribution of wages dominated by a core-periphery structure. This paper proposes that ϕ is high among communes within Chile, using three arguments to support this claim. First, the literature supports the existence of a core-periphery structure in Chile (Echeverria and Gopinath, 2007,

In the setup of the model, these are agricultural workers. However, in real life they can be considered as workers who cannot move because of economic, cultural or other reasons. They demand industrial goods, therefore they present an incentive for firms to stay in the southern region.

Aroca, 2009). Second, imports and exports among communes do not have costs, increasing the level of ϕ. This implies that ϕ within the country is pushed toward 1. Finally, migration of firms and workers is free of transaction costs, decreasing frictions for trade. I argue that the spatial distribution of wages is observed because agglomeration forces are stronger than dispersion forces. The empirical literature has focused on agglomeration and dispersion forces that can be measured, namely the market-access and housing-market effects. Details about the forces considered explicitly in the estimation process are discussed in the next section.

The Figure 2.3 represents the well known bell-shaped curve of spatial development that describes the situation for developing and developed countries Puga (1999). Starting from the left, low levels of ϕ (high transport costs) imply that dispersion is an efficient equilibrium, spatial concentration being the optimal result. High transport costs are present in developing countries, where high levels of concentration are an accepted stylized fact Puga (2002). High levels of ϕ (low transport costs) imply that concentration is no longer an efficient equilibrium, dispersion being the optimal result. This result is common for developed countries with multi-centered urban systems. I propose a simple method to determine where Chile is located in this picture. While Sudekum's model represents the first part of this picture, Helpman's model implies the reality for developed countries; therefore the simultaneous estimation of both models can provide information to evaluate the particular position of Chile on the bell-shaped curve of spatial development. To carry out this goal, I estimate Sudekum's model to represent the reality for developing countries. This simple test can generate a new perspective to classify developing and developed countries from a spatial perspective.

2.4 Model

The literature has considered the roles of the market-access and housing-price effects through estimation of Helpman's model (Brakman et al., 2004, Mion, 2004, Kiso, 2005, Hanson, 2005). However, the literature has not discussed the role of the housing-price effect under alternative extensions. I address this gap by estimating Sudekum's model, as well as Helpman's model. Using both models, this paper establishes whether or not the market-access and housing-price effects are relevant forces using different functional forms.

Similarly, Helpman's and Sudekum's models consider the role of the market-access effect, but they differ in the way the housing-price effect is incorporated.[11] Each time the models differ in a variable, a superscript h (Helpman) or s (Sudekum) is added. The consumers located in the spatial unit j have identical preferences defined by a well behaved Cobb-Douglas utility function:

[11] General derivations are specified in this section and more details can be obtained in the Appendix available in https://sites.google.com/a/ucn.cl/dusanparedes/.

$$U_j^h = M_j^{\mu} H_j^{1-\mu}, \qquad U_j^s = M_j^{\mu} H_j^{\gamma} A_j^{1-\mu-\gamma}, \tag{2.1}$$

where M_j represents the consumption of manufactured goods, H_j are units of nontradable domestically-produced goods and A_j is the standard agricultural good. The parameter μ represents the share spent on manufactured goods and γ represents an equivalent measure for domestically-produced goods. The agricultural sector is the numerarie good ($p^A = 1$), which is characterized by a competitive market with immobile workers, while the manufacturing sector follows a monopolistic competition model[12] (Dixit and Stiglitz, 1977). M_j is a composite good represented by a symmetrical CES function:

$$M_j = \left(\sum_{i=1}^{R} \int_0^{n_i} m_{ij}(z)^{\rho} \, dz \right)^{\frac{1}{\rho}}, \tag{2.2}$$

where ρ is the parameter of substitution among z varieties of the industrial good and m_{ij} is the manufactured good consumed in (j), but produced in (i)[13]. The imports in (j) can be obtained from R spatial units, where each region has a number of varieties equal to n_i. The Marshallian demand of the consumer located in (j) for a good produced in (i) is specified by:

$$m_{ij} = p_{ij}^{-\sigma} \mu Y_j G_j^{\sigma-1}, \qquad G_j = \left(\sum_{i=1}^{R} n_i p_{ij}^{1-\sigma} \right)^{\frac{1}{1-\sigma}} \tag{2.3}$$

where p_{ij} is the price of the manufactured good consumed in location (j), but produced in location (i), Y_j represents the consumer purchasing power (CPP) of the market located in (j), G_j is the composite price index for the industrial goods of location (j) and σ is the constant elasticity of the CES function, that is $\sigma = \frac{1}{(1-\rho)} > 1$.

The price differential for industrial goods between the origin (i) and destination (j) is attributed to transport costs, which are modeled using the iceberg transport costs proposed by Samuelson (1952). This type of transport cost implies that only a fraction of manufactured goods reach their final destination. If location (j) sends one unit of the industrial good, just $1/T_{ij}$ are received in location (j). If T_{ij} is considered the transport costs, price p_{ij} can be defined as:

$$p_{ij} = T_{ij} p_i, \quad T_{ij} = 1 \quad \forall \quad i = j \tag{2.4}$$

[12]In the original model, Krugman (1991) used the terms agricultural and industrial workers to refer to mobile and immobile workers. He established that a portion of workers in the southern region is not able to move because of linkage to assets generated by agricultural activity. However, in real life this group can be assumed as workers who cannot migrate to agglomerations. From a strict perspective, it is dispersion force, but not all textbooks include it.

[13]The core of the monopolistic model is represented by the effect of variety in the utility function. In contrast to competitive markets, the consumer prefers diversity. If $\rho < 0$, the varieties are complementary. If $\rho = 1$, variety as such does not matter. Finally, $\rho < 1$ ensures that the varieties are imperfect substitutes.

Using the equilibrium conditions, the wage equation can be specified by:

$$w_i = \left(\sum_{j=1}^{R} Y_j T_{ij}^{1-\sigma} G_j^{\sigma-1} \right)^{\frac{1}{\sigma}} \tag{2.5}$$

Equation (2.5) shows the role played by the forces described in the Section 2.3. To identify the market-access effect, the equation (2.5) can be redefined as:

$$w_i = \left(\frac{Y_i}{G_i^{1-\sigma}} + \sum_{j \neq i}^{R} \frac{Y_j T_{ij}^{1-\sigma}}{G_j^{1-\sigma}} \right)^{\frac{1}{\sigma}} \tag{2.6}$$

The market-access effect is composed of two parts: direct and indirect effects. w_i is a direct and positive function of the market-access effect, represented by real CPP (Y_i). Simultaneously, w_i is an indirect and positive function of market-access on surrounding locations determined by transport costs[14]. However, there are two problems with equation (2.6): the housing-price effect is missing in the specification and the estimation is not possible because the real data of G_j are not available. To overcome these problems, Hanson (2005) argues that, while nominal wages can be spatially different, real wages equalize labor markets among spatial units through inter-regional migration. If this condition is fulfilled, real wages are defined as:

$$\varpi = \frac{w_i}{\psi_i} = \frac{w_j}{\psi_j}, \tag{2.7}$$

where ϖ is a constant real wage and ψ_i and ψ_j represent the COL for location (i) and (j), respectively. The COL is the aggregated prices of agricultural $\left(P^A = 1\right)$, manufactured good (G) and housing (p^H). Helpman's and Sudekum's models incorporate housing prices through different specifications of the utility function (See equation (2.1)). Using the definition of the price index for the Cobb-Douglas utility function, the respective COLs for the two models are defined by:

$$\psi_i^h = G_i^{\mu} \left(p_i^H\right)^{(1-\mu)}, \qquad \psi_i^s = G_i^{\mu} \left(p_i^H\right)^{\gamma} \tag{2.8}$$

Inserting (2.8) into equation (2.7), the COL is specified as a function of w and p^H, and a constant ϖ. Inserting this price index G_i into equation (2.5), applying logarithms and adding a well-behaved error term ϵ_i; the wage equations to be estimated are specified as follows:

[14]This result represents the geography law in NEG: "*Everything is related to everything else, but near things are more related than distant things.*", which represents the geographical dimension incorporated by the NEG (Tobler, 1970).

$$\ln\left(w_i^h\right) = k^{'} + \frac{1}{\sigma}\ln\left(\frac{Y_i w_i^{\frac{\sigma-1}{\mu}}}{\left(p_i^H\right)^{\frac{(1-\mu)(1-\sigma)}{\mu}}} + \sum_{j\neq 1}^{R}\frac{Y_j w_j^{\frac{\sigma-1}{\mu}} T_{ij}^{1-\sigma}}{\left(p_j^H\right)^{\frac{(1-\mu)(1-\sigma)}{\mu}}}\right) + \epsilon_i \tag{2.9}$$

$$\ln\left(w_i^s\right) = k^{'} + \frac{1}{\sigma}\ln\left(\frac{Y_i w_i^{\frac{\sigma-1}{\mu}}}{\left(p_i^H\right)^{\frac{\gamma(1-\sigma)}{\mu}}} + \sum_{j\neq 1}^{R}\frac{Y_j w_j^{\frac{\sigma-1}{\mu}} T_{ij}^{1-\sigma}}{\left(p_j^H\right)^{\frac{\gamma(1-\sigma)}{\mu}}}\right) + \epsilon_i, \tag{2.10}$$

where $k^{'}$ represents a constant. Now equations (2.9) and (2.10) represent the direct and indirect effects of the market-access and housing-price effects. The market-access effect has already been explained, so I will focus on the housing-price effect. w_i is a direct and negative function of housing prices p_i^H. Simultaneously, w_i is a negative and indirect function of the housing-price effect of other locations p_j^H.

The NEG represents a spatial distribution of wages if the estimated parameters imply that agglomeration forces outweigh dispersion forces (Redding, 2010). If this is the case, the market-access effect is stronger than the housing-price effect, implying a spatial distribution of wages. These conditions are specified in Table 2.1. σ represents the nature of imperfect substitution among industrial goods, making the value of this parameter greater than 1. μ and γ are expenditures shares, therefore they must lie between 0 and 1. The parameter τ is crucial to evaluate the role of transport costs (See section 2.5). If $\tau > 0$, the distance is a correct variable to proxy transport costs. Moreover, $\tau > 0$ warrants that the indirect effect described in the equations (2.9) and (2.10) exists.

Table 2.1: Theoretical constraint of parameters

	Helpman	Sudekum
$\sigma > 1$	Substitution elasticity	Substitution elasticity
$0 < \mu < 1$	Share of industrial goods	Share of industrial goods
$\tau > 0$	Unit transportation cost	Unit transportation cost
$0 < \gamma < 1$	-	Share of housing
$\sigma/\left(\sigma-1\right) > 1$	Imperfect competition	Imperfect competition
$\sigma\left(\mu-1\right) < 1$	No-black-hole condition	-
$\sigma\left(\mu-1\right) > 1$	-	No-black-hole condition

Additionally, two relationships among parameters should be considered. The first, $\sigma/\left(\sigma-1\right) > 1$, reflects the market power given by imperfect competition, which is crucial to support the market-access effect[15]. The second relationship is derived from the the nonlinear equation system and is the called "no-black-hole"

[15]See the Appendix, the condition of marginal benefit equal to marginal cost is $p_i\left(\frac{\sigma-1}{\sigma}\right) = \beta w_i$. Imperfect competition implies that $\sigma/\left(\sigma-1\right) > 1$, otherwise there is no market power and the competitive market structure applies.

condition. This condition ensures the existence of the break-even point ϕ_B in a possible range of transport costs (See Figure 2.2). If the "no-black-hole" condition is not fulfilled, then *"the forces working toward agglomeration (regional divergence) always prevail and the economy tends to collapse to a point"* (Fujita et al., 1999). Following Brakman et al. (2009) the no-black-hole condition is specified by $\sigma(\mu-1)<1$ and $\sigma(\mu-1)>1$ for Helpman's and Sudekum's models, respectively[16].

2.5 Data, proxies and econometric issues

The estimation of (2.9) and (2.10) requires information about CPP (Y), transport cost (T), wages (w) and housing prices (p^H). The variables Y, w and p^H were obtained from the Chilean Household Survey CASEN 2006, which was conducted by the Ministry of Planning (MIDEPLAN). CASEN 2006 is a survey with representation at the national, regional, communal, urban and rural levels.

I estimated equations (2.9) and (2.10) with 334 communes, that represent the lowest level of spatial unit available in the CASEN 2006. This choice implies that Y, w and p must be aggregated from individual observations at the communal level[17]. With respect to Y, Brakman et al. (2004), Brakman et al. (2009) and Hanson (2005) use GRP, but this information is not available at the communal level in Chile. However, CASEN 2006 contains information about disposable income for each observation in a commune. Following the strategy of Mion (2004), I proxy Y as the weighted sum of disposable income for a sample of observations in a commune[18]. I argue that this measure better represents the CPP than GRP because it is the real income that people spend on goods, instead of GRP, which includes money spent outside the commune. Table 2.2 shows the descriptive statistics for CPP 2006. The correlation coefficient between wages and CPP is significant and positive (0.576), showing the relationship between the market-access effect and wages. The Moran's test (0.158) shows that CPP is positively correlated across space. This implies that rich communes are surrounded by rich communes and poor communes surrounded by poor communes, supporting the hypothesis that economic growth is spatially clustered.

Monetary information for T_{ij} is not available. Following Hanson (2005) and Mion (2004), an exponential decay and power distance function can be defined as proxies:

[16]For Helpman (1998) the sign is reversed because the interpretation of a reduction in transport cost is inverse: A reduction in transport cost increases dispersion. For details about this change, see Brakman et al. (2009).

[17]While Chile has 346 communes, only 335 are statistically represented in CASEN 2006. The ArcGIS map presented a problem with the commune called "La Higuera", therefore this paper considers 334 communes for analysis.

[18]Each observation of CASEN 2006 has a weight to expand the sample to the population. Using these weights, this analysis is expanded to the population level.

Table 2.2: Summary statistics

Variable	Mean	Std. Dev.	N
Wage	243118	43933	334
Consumer Purchasing Power (CPP)	5839	13589	334
IHPI	71734	45865	334
Sex	0.494	0.019	334
Education	8.819	1.401	334
Share highly skilled workers	0.122	0.092	334
Share fully employed workers	0.893	0.048	334
Share agricultural workers	0.315	0.214	334
Share mining workers	0.022	0.053	334
Share industrial workers	0.108	0.06	334
Share construction workers	0.084	0.044	334
Share service workers	0.036	0.037	334
Wage 2003	233077	41379	334
Correlation coefficients			
	Wage	CPP	IHPI
Wage	1.000		
CPP	0.576	1.000	
IHPI	0.595	0.528	1.000
Moran's test			
	Wage	CPP	IHPI
Wage	0.142*		
CPP		0.158*	
IHPI			0.156*

*Significant at 1% level.

1) All variables are estimated for 2006 unless indicated.

2) Wage and IHPI are expressed in Chilean pesos for 2006.

3) CPP is expressed in millions of Chilean pesos for 2006.

4) Sex is 0 for woman and 1 for man.

5) Education is the average of years for all workers at the commune level.

6) Moran's test was computed using the inverse distance matrix.

$$T_{ij} = \begin{cases} \exp(\tau d_{ij}) & \text{Specification 1 (Hanson, 2005)} \\ d_{ij}^{\tau} & \text{Specification 2 (Mion, 2004)} \end{cases}$$

where d_{ij} represents the geodesic distance (in kilometers) between the location (i) and (j), obtained from ArcGIS[19]. The parameter τ represents the gradient of reduction in transport cost according to distance; therefore $\tau > 0$. Geodesic distance means that $d_{ij} = 0$ when $i = j$, but the assumption of zero transport cost within a spatial unit is not realistic. I follow the recommendation of Head and Mayer (2000) by specifying $d_{ii} = 0.667\sqrt{\frac{area}{\pi}}$. This implies that distance within a spatial unit depends on its size.

The choice of the proxy w is crucial to represent the causal-mechanism described in section 2.3. Following the insights specified by Helpman (1998) and Sudekum (2006), w represents the level of attractiveness of a location to encourage the migration of workers and firms. The literature recognizes that workers with low education tend not to migrate due to the high cost involved, therefore they do not look at the complete distribution of wages in making the migration decision[20]. To aggregate wages at the communal level, I calculated the average wage of the economically active population (EAP) considering the wages between 25% and 99% of the wage distribution of each commune. The upper boundary was choosen to reduce the impact of outliers with extremely high wages.

Figure 2.2 presents the core-periphery structure of wages[21]. The mining and salmon clusters are clearly identified in the extremes of the country, but the map does not show the concentration of high wages around the MP so a zoom map is needed. In the zoom map, the MR reveals a core-periphery pattern that is supported by the Moran's test result of 0.189. This value indicates that the spatial relationship of w is positive, with high wages surrounded by high wages. Finally the spatial autocorrelation of wages is greater than any other variables, indicating that wages are strongly clustered, supporting the NEG analysis.

Following Brakman et al. (2004) and Kiso (2005), I represent p^H as the average housing price for each commune. Table 2.2 indicates that p^H correlates strongly with higher wages (0.595), suggesting that communes with higher wages also have higher housing prices. Simultaneously, p^H has a high spatial autocorrelation (0.156), suggesting that communes with high p^H are surrounded with communes with high p^H. This sup-

[19]Transport cost is better represented by average car travel time between locations, however this information is not available. Geodesic distance can be assumed as the flight-distance between two spatial units.

[20]This consideration is equivalent to the third dispersion force proposed by Krugman (1991), which is not discussed by Baldwin et al. (2003)

[21]Shape file of the map of Chile can be downloaded free of charge from http://www.diva-gis.org/. The map shows the 334 communes with statistical representativity, leaving empty spaces for the communes without statistical representativity in CASEN 2006.

ports the evidence that the market-access effect is spatially clustered. In summary, descriptive statistics appear to support the theoretical relationship among wages, income and housing prices.

Finally, the estimation of equations (2.9) and (2.10) is affected by endogeneity. The first source of endogeneity is measurement-error in housing prices and CPP. The careful choice of proxy variables for p^H and Y should reduce this bias. The second source of endogeneity is related to the variable w, which is present on both sides of the wage-equations, implying that $E[w_j|\epsilon_i]\neq 0$ for $j = 1 \dots R$ each time that wages are evaluated in the same year. However, the wage of a location j in 2006 (w_j^{06}) can be broken down into a variation from 2006 (Δ^{06}) and a variation from 2003 (Δ^{03}). Technically, the main source of endogeneity is derived from (Δ^{06}), but (Δ^{03}) must be uncorrelated with the (w_j^{06}). Thus, using the predicted values $\hat{w}_j^{06}$ derived from the regression between w^{2006} and w^{2003} should reduce endogeneity bias.

Table 2.3: Fitted wage

	(1) Wage 2006
Constant	49432.12
	(0.000)
Wage 2003	0.83
	(0.000)
N	334
R^2	0.61

1) p-values in parentheses

2) p-values are based on White's heteroscedasticity-adjusted standard errors.

3) Estimation with OLS.

A limitation of this methodology is the availability of wage data at the communal level for 2003. Using the CASEN 2003, wages were estimated using the same procedure described for CASEN 2006. However, CASEN 2003 has a statistical representation of 302 communes; so imputation is needed for the 32 remaining communes. In the communes without information for 2003, I replaced the fitted values with the w of 2006. Another way to deal with endogeneity problems is to use GMM for the estimation, but the functional forms of equations (2.9) and (2.10) make this process extremely complicated. I estimated equations (2.9) and (2.10) using GMM, but the logarithmic functional form implies a flat objective function and convergence was not achieved. However, the procedure that I present with the fitted value of wages must reduce the inherent problems of endogeneity attributed to NEG models. In summary, to estimate equations (2.9) and (2.10), the wages on the right-hand side were replaced by the predicted value derived from the regression detailed in table 2.3. To control for endogeneity, a non-linear estimation was carried out using Stata software[22].

[22]The codes and data are available upon request.

2.6 Results

The results of the estimation of equations (2.9) and (2.10) are shown in table 2.4. Columns 1 and 2 contain the structural parameters for Helpman's model under both specifications of transport costs defined in equation (2.11). Comparing the structural parameters in Table 2.1, all of them are significant and have the expected sign . In Helpman's model, the value $(1-\mu)$ represents the expenditure of household goods. According the the National Expenditure Survey 1996, the expenditure on household goods is around 14% of the total income, which is close to the value estimated (15.9%). A small change in μ is observed for the distance functional form, but still highly consistent with theoretical requirements (13.9%).

A key parameter is the elasticity of substitution σ, the estimation of which is significantly greater than the one for Helpman's model. This implies a profit margin of 2.3% attributed to imperfect competition. In spite of the correct sign of σ, the profits derived from imperfect competition are low compared to estimations of the US (14%) and Germany (38%) (See Hanson (2005) and Brakman et al. (2004) respectively). The profit margin does not change for different specifications of transport costs, supporting the consistency of its estimation. As discussed in section 2.4, the no-black-hole condition is another relevant condition, which is represented by $\sigma(1-\mu) < 1$. For Helpman's model, the no-black-hole condition is rejected for both transport costs, implying that agglomeration is a stable equilibrium, independent of the level of transport costs. Given that the no-black-hole condition is rejected, any comment about τ is meaningless. In summary, Helpman's model shows parameters that support the spatial distribution of wages in Chile, but reflects a reduced profit margin derived from imperfect competition.

The conclusions are slightly different for Sudekum's model (column 3 and 4). The parameter γ, namely the share of housing, is estimated at around 16% for both transport costs. The parameter μ does not change considerably when using an exponential specification, but it increases to 0.88 for the distance function. The parameter σ is very similar to that in Helpman's model, implying a similarly low margin profit of 2.3%. The no-black-hole condition is fulfilled for both specifications of transport cost, indicating that an increase in the level of freeness among communes implies a higher heterogeneity of wages. Columns 3 and 4 show that τ is significant and positive, suggesting that the market-access effect has a positive impact on wages, but is greater when locations are closer. In summary, both models suggest a core-periphery for Chile with a market-access effect that is stronger than the housing price-effect. However, the level of freenees is only relevant for Sudekum's model.

As discussed in the previous sections, the case of developing countries is a mix between the skill composition of workforce and first and second nature. To control for skill and first nature, several variables are added, such as the number of workers participating in the mining sector, average education and the proportion of highly skilled workers, as proposed in Brakman et al. (2004). The incorporation of these controls variables are shown in the columns 5, 6, 7 and 8 of the table 2.4. The set of control variables are described in table 2.2, but just the significant ones are reported in table 2.4. They have the expected signs. For example, the proportion of male workers positively affects the local wage, supporting the gender gap in wages. The relationship between the proportion of highly skilled workers and wages is positive and significant. Finally, the proportion of agricultural workers has a negative impact on local wages, representing the negative effect of primary activities. The incorporation of these variables implies some changes in the parameters, especially in σ which loses its significance every time that the power distance function is used. Additionally, the σ coefficient is unstable for different initial values under the distance function, reducing the stability of this functional form. The p-value goes up for the power distance function specification in both models, reducing its level of fitness. Using this result, I choose to keep the analysis with the more stable specification derived from the exponential distance function.

Table 2.4: Structural estimation

	Helpman (1)		Sudekum (2)		Helpman (3)		Sudekum (4)		Final model (5)	
	Exponential	Distance	Exponential	Distance	Exponential	Distance	Exponential	Distance	Sudekum	Helpman
σ	44.96	46.01	44.96	47.38	40.44	41.80	33.57	40.54	42.14	41.06
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.045)	(0.000)	(0.854)	(0.001)	(0.000)
μ	0.841	0.861	0.841	0.887	0.754	0.780	0.623	0.756	0.780	0.759
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
τ	0.103	0.029	0.104	0.061	0.000206	0.0135	0.00267	0.0171	0.003	0.001
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.002)	(0.000)	(0.014)	(0.000)	(0.000)
γ			0.159	0.158			0.429	0.310		0.308
			(0.000)	(0.000)			(0.000)	(0.328)		(0.000)
$\sigma/(\sigma-1)$	1.023	1.022	1.023	1.022	1.025	1.025	1.031	1.025	1.024	1.025
$\sigma(1-\mu)$	7.149	6.395	7.149	5.354	9.95	9.20	12.66	9.89	9.271	9.900
Wald test			0.25	1.33			17.69	0.32		18.90
$(1-\mu)=\gamma$			(0.620)	(0.000)			(0.000)	(0.572)		(0.000)
Sex					0.107	0.471	0.346	0.635	0.370	1.015
					(0.663)	(0.100)	(0.317)	(0.036)	(0.274)	(0.005)
High skill					1.655	0.895	1.088	1.436	1.333	1.316
					(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Full time					-0.325	-0.300	-0.104	-0.0465		
					(0.003)	(0.006)	(0.427)	(0.687)		
Agriculture					-0.376	-0.390	-0.298	-0.537	-0.394	-0.425
					(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$\sqrt{MSE}$	0.444	0.911	0.446	0.176	0.457	0.790	0.347	0.121	0.448	0.221
N	334	334	334	334	334	334	334	334	334	334

1) p-values in parentheses.
2) p-values are based upon White's heteroscedasticity-adjusted standard errors.
3) Estimation with NLS.
4) The coefficients for the rest of the control variables are available upon request to the author.

In spite of controlling for several sources of agglomeration, the concern about endogeneity still remains. Using the methodology discussed in section 2.5, the equations (2.9) and (2.10) are estimated using the fitted variable $\hat{w}_j^{2006}$ in the right-hand side of the equation. As Table 2.4 suggests, endogeneity generates an under-estimation of σ. The parameter σ increases significantly for both models, indicating a higher imperfect substitution of manufactured goods. μ and γ are within the theoretical requirements. Columns 9 and 10 of Table 2.4 show the importance of the neighborhood to determine the wage level through the positive value of τ. Locations with high wages are surrounded by locations that also have high wages, supporting the earlier results suggested by the Moran's test. This parameter is very stable, independent of the model estimated. Control variables have the right sign. The profit rule is fixed around 2.3% for both models, confirming the presence of imperfect competition. However, only Sudekum's model supports the no-black-hole condition, suggesting that the core-periphery structure can increase even for a continuous reduction of transport costs. Helpman's model does not fulfill the no-black-hole condition, indicating that total agglomeration is a stable equilibrium independently of dispersion forces. Although comparison of the models is not the main goal of this paper, the Wald test suggests that the parameter γ is significantly different from $(1 - \mu)$, arguing that both models explain the core-periphery structure. However, only Sudekum's model fulfills the no-black-hole condition and this model should represent more properly the spatial distribution of wages in Chile. This result implies that Chile seems to be located on the right side of the bell-shaped curve of spatial development, fitting with the case of developing countries and supporting the empirical literature (Echeverria and Gopinath, 2007, Aroca, 2009).

2.7 Conclusions

The NEG uses the combination of increasing returns, transport costs and imperfect competition to explain the spatial distribution of wages. In this paper I suggest that Chile fits with the predictions derived from the NEG, but with weak evidence of imperfect competition. The spatial distribution of wages is mainly explained by the market-access effect being stronger than the housing-price effect. This implies that the MR, even with congestion costs derived from agglomeration generates higher agglomeration economies with respect to the other regions in Chile. This result is supported by the estimation of the models proposed by Helpman (1998) and Sudekum (2006). Simultaneously, the empirical evidence suggests that Sudekum (2006) fulfills the no-black-hole condition, suggesting the case of Chile as a developing country. From the policy perspective, this suggests there will be higher levels of wage disparities in the future because agglomeration is a stable equilibrium. These results are in line with previous research of Aroca and Bosch (2000), Echeverria and Gopinath (2007) and Aroca (2009).

With respect to the structural estimation of Helpman's and Sudekum's models, both are theoretically consistent for almost the complete set of theoretical requirements. The coefficients σ, μ, γ and τ lie in the theoretical ranges specified by the NEG model. This consistency is robust for two functional forms of transportation costs, namely the exponential distance function and the power distance function. The shares μ, γ are precisely estimated according to the evidence provided by the National Expenditure Survey 1996. However, the control of the skill composition of the workforce and first nature do not appear to modify the estimations for both models. This could be explained by the high levels of nonlinearity of equations (2.9) and (2.10), where the ad hoc incorporation of a linear set of variables has a marginal impact on an extremely flat function. These control variables only changed the significance of the parameter σ, specifically for the power distance functional form of transport costs. This result is in line with the literature, where the exponential decay functional form has been widely used in empirical estimations. Considering the exponential decay function, both models show precise estimations for the structural parameters.

A final adjustment was the control of endogeneity. Using the fitted values $\hat{w}_j^{2006}$, the structural parameters are consistent for both models. However, the no-black-hole condition is rejected for Helpman's model. This implies that concentration of wages around the MR is a fixed result, independent of the level of freeness trade. However, Sudekum's model satisfies the no-black-hole condition, providing evidence about the role played by the level of freeness of trade. This opens a new empirical alternative to evaluate the market-access and housing-price effects. Sudekum's model was estimated for the first time and its performance was suitable for the case of Chile, providing evidence about the spatial distribution of wages at the communal level. This implies that Chile should be located on the left side of the bell shape. With this evidence, higher levels of agglomeration can be expected, consistent with a developing country.

In spite of the theoretical requirements, intuition must play a role in interpreting the results derived from the econometric exercise. The core-periphery structure rests on the crucial assumption of imperfect competition. For the complete set of models, the power market derived from imperfect competition implies only a 2.3% margin on the competitive equilibrium price. This is extremely low with respect to the literature and the evidence must be considered carefully. This result can be related directly to the combination of characteristics present in developing countries. Ad hoc incorporation at the aggregated level (communal) does not seem to properly capture the additional variation of wages. New directions are needed for the special case of developing countries. One is the estimation of the models using microdata, where the controls of skill composition and first nature can be incorporated in the estimation of the causal mechanism. Future research in this area will provide information on where Chile is located on the bell-shaped curve of spatial development, as well as the role played by second nature in the spatial distribution of wages.

Chapter 3

Spatial wage inequality: a multilevel competition among human capital, NEG and amenities.

Road map for Chapter 3: The Section 3.2 relates the case of Chile with a particular developing country. This section discusses the particular role played by the natural resources on the economic geography of Chile. Section 3.3 shows two theoretical models: NEG and SEM. The section 3.4 gives a detailed idea about the econometric strategy for both models. The Section 3.5 shows the varaible obtained from the new data set CASEN 2009. Finally, Sections 3.6 discusses the results. They suggest that NEG is not sufficient for explaining the spatial wage inequality in Chile and the role of natural resources is crucial for evaluating the Chilean economic geography.

3.1 Introduction

Spatial wage disparities are a constant concern for policy makers, as well as the subject of popular debates. A clear example is the European Commission, which has supported several policies to reduce regional disparities through direct investment at the inter- and intra- country level. In Latin American (LA) countries, significantly higher levels of inequality have led to strong recommendations from international groups such as the ONU-WIDER that these countries focus on this problem since increasing inequalities have arisen "*partly as a consequence of the uneven impact of trade openness and globalization*" (Kanbur and Venables, 2005). Accurate conceptualization and empirical estimation of the underlying fundamentals driving the observed disparities are necessary both to design, and to assess the effectiveness of, such (proposed) regional equalization polices.

The literature has addressed the inequalities from a number of different perspectives and at both micro and macro/regional levels. The most developed approach is Human Capital Theory (HCT) where the wage differential is market driven by demand and supply of different human capital endowments and the level of analysis is the individual worker. HCT suggests that highly skilled workers are concentrated in large urban areas, which explains the wage inequality in core-periphery structures, also referred to as spatial sorting of skills. Combes et al. (2008a) test this phenomenon and find that half of wage inequality is explained by

HCT. In this sense, the HCT is the main framework for analyzing the wage differentials and its empirical performance is hardly questionable. However, HCT focuses on the micro level of the worker, while others forces operate at a more aggregated level such as the geographical scope of the labor market (Rosenthal and Strange, 2004). A second group of approaches to examining wage inequalities, mainly in regional economics, is conducted at a geographical scale rather than at the level of the individual. Within the spatial/geographic analyses at least two main approaches compete for explaining the wage variation that is not explained by HCT alone.

The first approach, commonly referred to as New Economic Geography (NEG) following Krugman (1991), implies a positive relation between wages and Market Access (MA)[1]. This framework indicates that the spatial concentration of firms decreases the average cost of production, providing the incentives for firms to seek agglomerations and the ability to offer higher wages, conditioned by a production function with increasing returns to scale. While the NEG has been shown to perform well empirically for several developed countries[2], it does not take into account the natural differences among regions, such as raw materials endowment and climate features (Combes et al., 2008a). This missing dimension poses a problem for explaining wage inequality in countries where, for example, the exploitation of natural resources is a key sector of the economy. In contrast to NEG, a second approach, the Spatial Equilibrium Model (SEM) (Roback, 1982) brings out amenities as the main explanation for wage disparities. Roback suggest that if the shadow prices of amenities are capitalized in wages and rents, then the spatial wage inequality should be reduced after controlling by both variables. These amenities are any initial endowment, such as sea access or climate, which affect the utility of consumers or cost function of the firm. Roback (1982, 1988) and Dumond et al. (1999) among other authors, show that, after controlling for the shadow price of amenities, the nominal wage inequality vanishes. While both theories have found empirical support in different countries, there is no doubt that NEG and SEM compete in explaining the local variations in labor supply and demand that cannot be explained by HC alone (see Partridge (2010), for a thorough discussion.).

Considering the theoretical approaches described above, how can the NEG and SEM theoretical frameworks be articulated for empirical investigations of inequalities in the case of Latin American countries where income (and related) inequalities are highly conditioned, among other factors, by economic geography? This paper discusses the main challenges derived from the particular context of Latin America. First, the transportation systems, market structures and institutions are at a very different stage of development compared with those in the USA and the European countries, for example. Secondly, LA countries show a

[1] According to Garcia (2006) the majority of the articles estimate an elasticity wage-MA between 13 and 30 percent

[2] For empirical applications see U.S. (Graves et al. (1999); Dumond et al. (1999); Hanson (2005); Partridge et al. (2009)), Germany (Brakman et al., 2004), Great Britain (Fingleton, 2006, 2009) and Italy (Mion, 2004)

high dependency on natural resources exports rather than manufacturing and service exports. Thirdly, high transport costs in LA, compared to developed countries, reduce the ability of market solutions based on factor mobility to reduce the spatial disparities. Undoubtedly, a large set of structural differences compared with more highly developed economies remains. Moreover, each one of these characteristics may affect the empirical performance of the theories described, especially those specified at an aggregate level such as NEG or SEM.

This paper contributes to understand a particular of a Latin American country, Chile, by proposing several theoretical and empirical innovations relative to the existing literature. First, Venables (2005) suggests that the spatial endowment of natural resources is crucial for understanding spatial location of economic activity in primary goods export oriented countries. According to the British Geological Survey (2008), Chile is the main global copper producer, thus a country with a high dependency of natural resources. To capture the natural resources role, this paper casts natural resources as a particular type of firm amenity in the SEM framework. That is, amenities are not represented only by standard proxies such as climate or quality of life as discussed by Graves (1976, 1979, 1980), but rather they are immobile productive resources. This representation allows the SEM to capture the particular context of natural resource dependence, in the competition between SEM and NEG for explaining the portion of spatial variation in wages beyond what HCT can account for.

A second set of innovations is proposed from the empirical perspective. The existing literature has estimated the explanation power of NEG and SEM models using strategies with several shortcomings. In the case of NEG estimations, Brakman et al. (2004, 2006), Breinlich (2006), Fingleton (2005), and Brulhart and Koenig (2006) among other authors have estimated a wage equation using aggregate level variables, namely average wages at different spatial scales. The main concern with this approach is the exclusion of the role of human capital such as the distribution of education level or experience. When micro-level information is unavailable, the aggregate approach is a valid one. However, ignoring individual information renders this approach at least problematic. Fally et al. (2010), Hering and Poncet (2010) are exceptions in estimating wage equations using micro data level. While these latter articles incorporate worker heterogeneity, they use as the dependent variable the MA which works at an aggregated spatial scale[3], thereby ignoring the interaction between both levels: micro and macro scales. For example, MA can affect individual wages through the scale economies enjoyed by geographically clustered firms (NEG proposition), but this concentration can also affect the quality of workers such as education or productivity. Ignoring the interaction between the micro and aggregate levels "*is dangerous at best and disastrous at worst*" (Atkin and Longford, 1986). This

[3]The MA is computed at county, province, state or even country level.

paper contributes a means of overcoming these problems associated with the consideration of a bi-level analysis-both micro and regional. As far as the author knows, this is the first paper using this approach to analyze spatial wage inequality. The bi-level analysis captures worker heterogeneity (micro level), while the NEG and SEM models are considered at a regional level (county). This methodology incorporates the interaction between both levels and the estimations should thus be seen as an improvement on the existing literature.

A third empirical contribution is directly related with Chilean economic geography. As shown in figure 3.1 and 3.2, the extreme north of the country is characterized by a production system based on the mining industry, specifically copper. The central regions are characterized by large concentrations of population, where the Metropolitan Region represents the 41% of the population and the tertiary sector supports the economy. Simultaneously, Chile extends 4,270 kilometers from north to south and 177 kilometers (on average) from east to west, a geographical shape that imposes special considerations on the determinants of wage differentials. First, the high transport costs at the intra-country level, hinders the clearing condition of regional labor markets and wage differentials that could appear (Aroca and Hewings, 2002).[4] Secondly, the spatial location of firms may be regarded as a macro representation of Hotelling's problem, where the reduction of transport costs leads the industry and workers to be located in the middle of the country (MR) (Echeverria and Gopinath, 2007) generating spillovers affecting worker productivity through every channel described by Marshall-Arrow-Romer externalities and, again, manifested in wage differentials. Finally, the industries located on both extremes of the country can offer high wages to compensate for the transport cost of working in those regions, but living in a place without the built amenities associated with a capital city such as the MR (Aroca and Atienza, 2008). All these factors suggest that the economic geography of Chile is not adequately represented by only one causal mechanism, namely MA or natural resource dependence, but rather Chile could be a hybrid case. In order to make a fair competition between NEG and SEM, this paper proposes an empirical exercise where the NEG framework can vary its performance across space. The a priori hypothesis is that this model is more appropriate in dense areas, such as the central area of the country, but its explanatory power decreases in those regions where the natural resource base is more prominent.

[4]The distance by road between Arica (the city located at extreme north) and Punta Arenas (the city located at extreme south) is around 5185 kilometers. This implies a driving time of 65 hours without any stop and with an average speed of 80 kilometers per hour.

Figure 3.1: Administrative division of Chile.

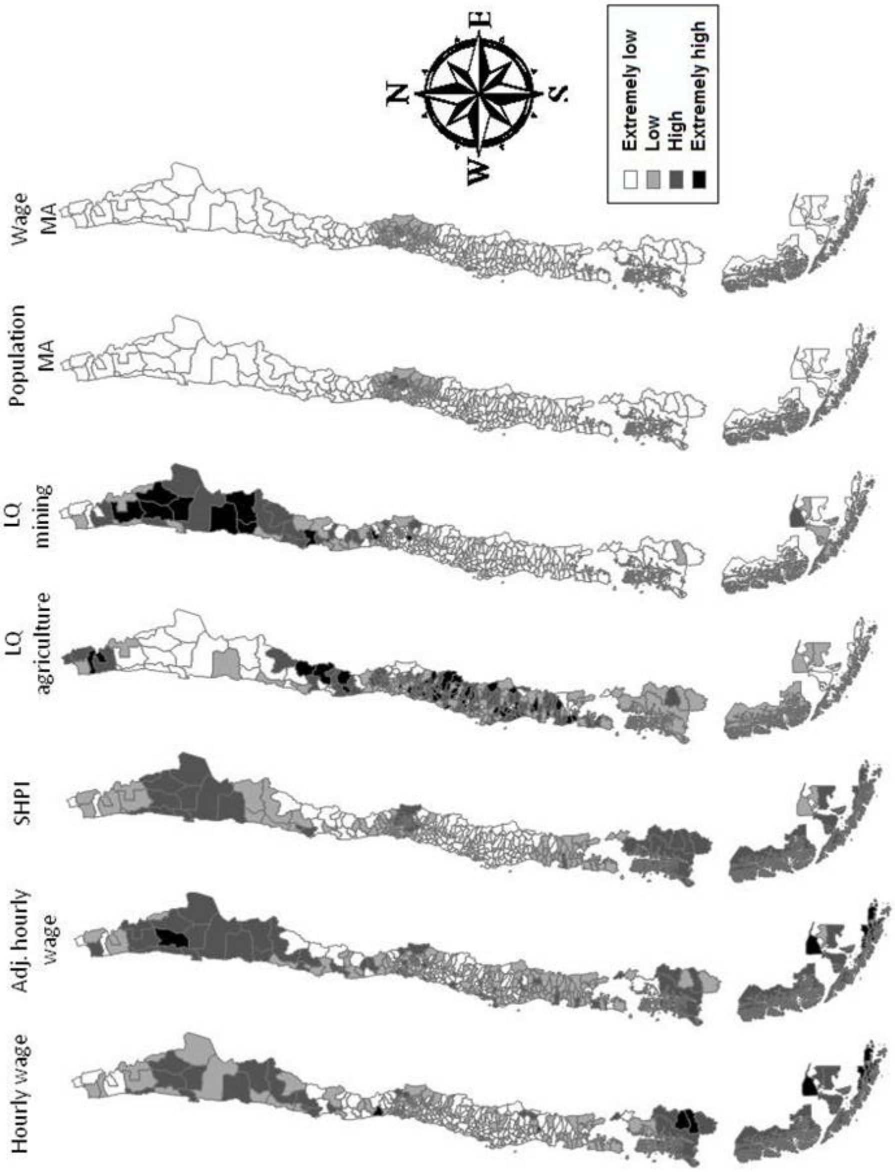

Figure 3.2: Economic Geography of Chile.

The empirical results confirm that the approach taken in this paper produces a richer and more nuanced understanding the Chilean case. The estimations show that 12% of the wage variance can be attributed to the community level, which is labeled as initial spatial wage inequality. Human Capital Theory exhibits the best performance reducing the initial spatial wage inequality by 53.85% and a 46,15% of spatial wage inequality is maintained. When average education and sex are added to HC, the spatial wage inequality is reduced by 70.52%. These estimations suggest that wage inequalities are strongly explained by different mixes of human capital endowment in the regions that may also be a reflection of differential economic opportunities. This result, as far the author knows, is the first estimate of the human capital and other spatial components of wage inequalities for the Chilean case. The rest of the spatial wage inequality, namely 29.48%, could be explained by both macro theories: NEG and SEM. The spatial wage inequality is poorly explained by the NEG. The MA-wage elasticity is extremely low in comparison with estimations in other countries and the NEG explains around the 6% of the initial spatial wage inequality. The interaction between HC and NEG does not perform much better than HC alone because both theories reduce the spatial wage inequality by a 70.56%. In contrast, the SEM performs much better. The incorporation of housing price, such as Roback (1982) suggest, accounts for 57.69% of initial spatial wage inequality with a housing price-wage elasticity of 0.61: if the wage increases at 10%, then housing prices are increased in a 6%. The role of natural resources, as a type of firm amenity appears to be a crucial factor and its explicit consideration reduces the wage differential to 78.02% when it is interacted with HCT. Finally, the MA performance is low for those regions where natural resources sectors are prominent, but it performs better in the densely populated regions. The results indicate that wage inequality is affected more by natural resource amenities than by MA differences, thus opening up new avenues of research regarding the role of NEG in the case of developing countries.

The paper is structured as it follows. Section 2 presents the Chilean context and how this case could represent developing countries in general. Section 3 presents the theoretical models, namely NEG and SEM, and describes their a priori fit for the Chilean case. The econometric identification strategy and data comprise Section 4, followed by results in Section 5 and conclusions in Section 6.

3.2 The Chilean context: an example of a developing country

Chile presents a particular economic geography that must be understood when evaluating the wage differentials. Kanbur and Venables (2005) and Aroca (2009) among other authors have highlighted the excessive spatial concentration of the economic activity around the capital of the country, namely the Metropolitan Region of Santiago (MR). According to Echeverria and Gopinath (2007), the MR occupies 2% of the total territory, but 50% of the total production and the 40% of the total population are concentrated in this area. Aroca (2009) also highlights the concentration of the institutional design around the MR, where an extreme centralized political structure dominates the autonomous development of the regions. These stylized facts undoubtedly determine the economic geography of the country, and obviously influence the wages.

The NEG literature links geographical concentration with wages. According to this theory, the size of the MA is positively correlated with wages. Looking at figure 3.2, this idea partially fits with the Chilean case. The MR presents a large MA where the firms prefer to locate and they compete with each other for skilled workers who are able to claim higher wages; however this variable is irrelevant for other geographical areas in Chile where the wages are also high such as the extreme north. Given this spatial pattern, the NEG appears as a natural model to evaluate the case of the MR, but not the general Chilean case.

Alternatively, the SEM suggests that amenities, coming from the consumer or producer side, simultaneously affect wages and rent prices. Those regions with consumer amenities compensate with the lower wages and those with dis-amenities generate premium wages. This paper assumes the natural resources as a particular type of amenity, which mainly affects the production side or producer amenity. The lower cost attracts firms generating a geographical concentration, and increasing the worker productivity through the economics of agglomeration (Krugman, 1991). This mechanism explains why spatial units with natural resources present higher wages. Simultaneously, the location of firms and workers pushes up land demand, and those regions with producer amenities could also present higher housing prices. To support the fit of the SEM theory on the Chilean case, figure 3.3 plots the wage differentials and the Spatial Housing Price Index (SHPI) for 2009.[5] The horizontal axis represents the spatial distribution of counties for Chile where the left corner represents the extreme north and the right corner the extreme south[6].

[5] This index is proposed by Iturra and Paredes (2011). Given that Chile does not estimate an official measure; this variable is the best available proxy to estimate the role of the housing prices.

[6] The wage differential is approached with counties fixed effects after controlling by characteristics of the worker and industries fixed effects.

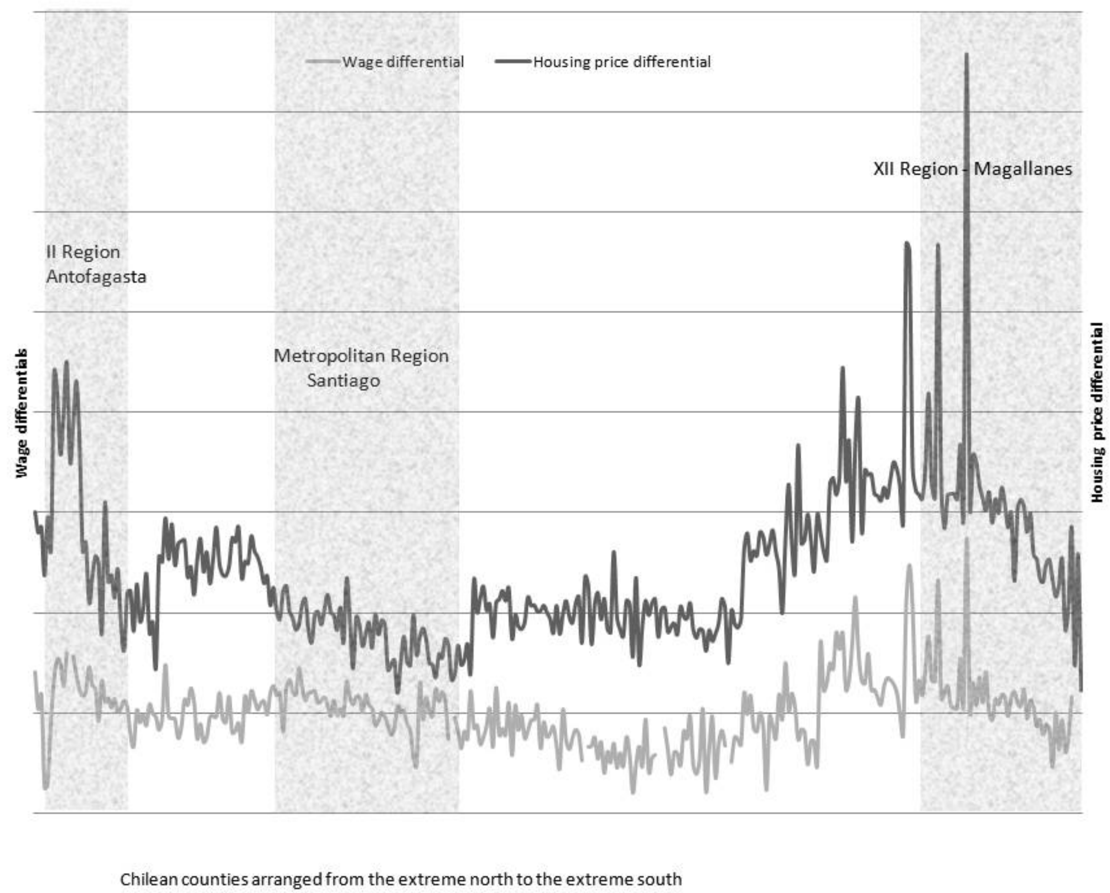

Figure 3.3: Spatial distribution of wages and housing price differentials, counties arranged north to south. Chilean map is "tended" on the horizontal axis.

According to figure 3.3, the highest level of wages is found around the Metropolitan Region of Santiago, which also represents the highest spatial concentration of population in the country. Two small clusters of higher wages are located in the extremes of the country. The II Region of Antofagasta and the extreme south also exhibit higher wages. The Spatial Housing Price Index follows a similar pattern, with the highest rents centered in the MR followed again by the extremes of the country. This plot supports the graphical representation of figure 3.2 as far as the outcomes of the regions with natural resources display higher wages and rents, though not enough to overwhelm the impact on the MR. According to the stylized facts described previously, the NEG and SEM frameworks seem to apply for the Chilean case. The next section explains the theoretical background of both models and how they can be represented in the Chilean context.

3.3 Theoretical model

This section discusses the theoretical aspects of both competing models: NEG and SEM. Each one of the models proceeds with an empirical explanation to support its use to evaluate the Chilean case. This exercise is only a descriptive analysis and must not be considered as a final evidence for evaluating both models.

3.3.1 Spatial Equilibrium Model (SEM)

Assume a set of cities where each one owns a level of amenity s which varies continuously over (S_1, S_2) and it is exogenously fixed. The workers-residents are producers and consumers of a composite consumption good X. Commuting among cities is prohibited, but the migration is allowed as well as the intra-city commuting. Workers have similar tastes and skills and they offer a single unit of labor. Thus, the worker maximizes the level of utility consuming a composite good X and residential land l^c. The market equilibrium condition for wages w and rents r given an amenity s is captured by the indirect utility function:

$$V(w, r; s) = k \tag{3.1}$$

The indirect utility function has the usual properties and $\partial V / \partial s > 0$ because s is an amenity. The indirect utility must be similar across space avoiding incentives for migration of workers. The composite good X is produced by firms with a constant-returns-to-scale production function $X = f(l^p, N; s)$ with l^p land used in production and N for the total number of workers and can be considered a measure of economies of agglomeration. Given the structure of the production function, the unit cost must be equal to the product price normalized to 1:

$$C(w, r; s) = 1 \tag{3.2}$$

This condition removes the incentive of firms to move among cities. The marginal costs are represented by $C_w = N/X$ and $C_r = l^p/X$. The amenity can affect the firm's plan through two channels. If the amenity is unproductive, then $C_s < 0$. The opposite sign applies when the amenity is productive. Equilibrium implies a w,r combination which depends on s. Thus, the magnitude of spatial variation in wages and rents in equilibrium is related to the amenity on either or both the consumer and firm side. This causal mechanism is described by the figure 3.4 which evaluates how the equilibrium condition applies for the case of two regions and for one representative consumer who lives and works in a spatial unit characterized by an s level of amenity. The consumer derives a utility level equal to k such as $V(w, r; s) = k$. Simultaneously, the firm also faces this level of amenity producing a unitary cost $C(w, r; s) = 1$. Both conditions imply an equilibrium combination of w_s, r_s described by the point A in both graphs.

Thus, the empirical version of the SEM model postulates that wages and price level differentials are related with the role of amenities. Using the Chilean case the consumer has two location choices: he can move to a city with a high level of natural resources (graph on the top in figure 3.4), say Region II (Antofagasta), characterized by an s_a level of amenity; or he can move to the Metropolitan Region of Santiago (MR) with an s_s level of amenity (graph on the bottom in figure 3.4). The amenity in Antofagasta is represented by the access to the copper mining. This amenity implies that the city is mainly characterized by industrial production and it is not considered a pleasant place for consumers by assumption, then $\partial V/\partial s_a < 0$. In order to keep the same level of utility, the amenity s_a implies that workers require a higher level of wages and/or lower level of rents (B) if they move to Antofagasta. In contrast, this amenity positively affects the production function of firms, $\partial C/\partial s_a > 0$, since it is a producer amenity. The firms can offer higher levels of wages and they must pay higher rents in competition for land with other firms, moving from B to C. These assumptions imply a high level of wages and housing prices in the city with natural resources. The increment in wages is unambiguous while the impact on rents depends on the relative sizes of the C and V shifts.

Simultaneously, consumers can also move to the MR; the city is characterized by a concentration of public goods, a high level of urban amenities such as up-scale shopping and entertainment and knowledge spillovers. These amenities are positively perceived by workers moving toward D and $\partial V/\partial s_s < 0$ by assumption. These amenities also affect the cost of the firms, which enjoy the benefit of agglomeration economies due of the existence of thick labor markets and backward or forward linkages (Marshall, 1890), generating a movement to E. As the reader can see, the movement from A-E (bottom) is larger than A-C (top). Note that in the

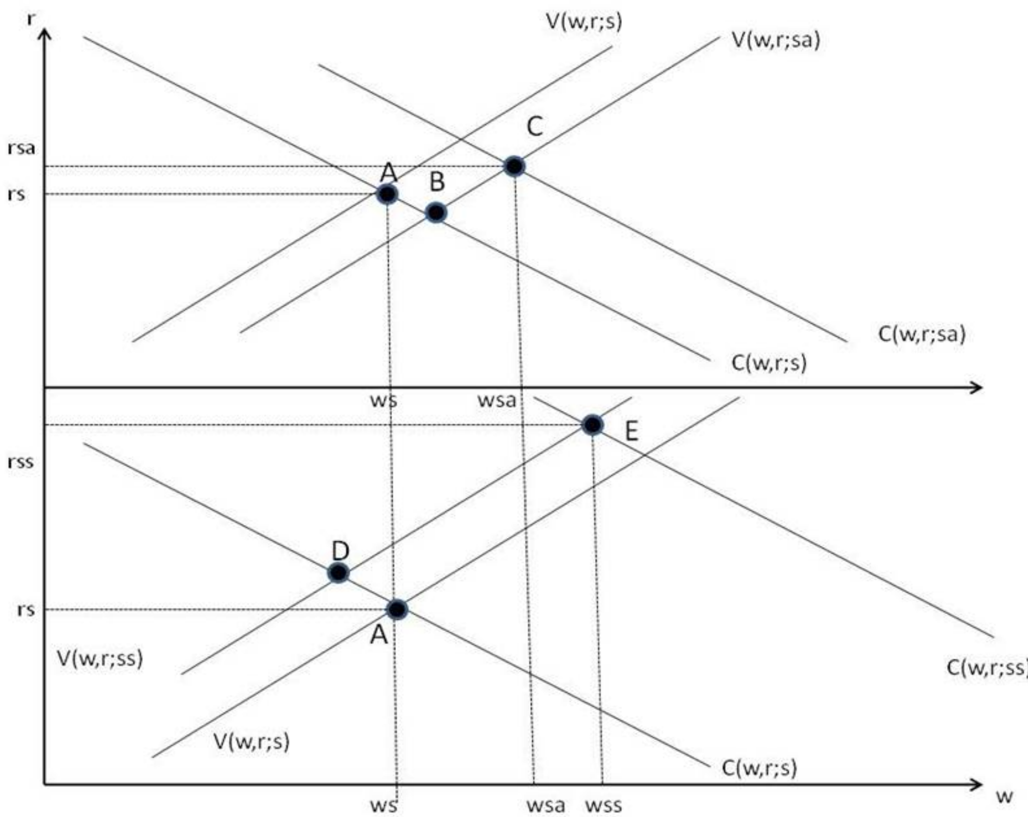

Figure 3.4: Explanation for the Chilean case using the Roback's (1982) model.

first case, the movement is generated by spillovers of the natural resources in the cost function. The main channel can be attributed to economies of specialization in the production of the natural resources outputs which increases the productivity of workers. However, the impact of amenities on the cost function in the case of MR is based exclusively on the urbanization economies, technological spillovers or learning process, which according to Brakman et al. (2005): "*agglomeration advantages lock business activity in relatively prosperous core regions, even though wages - and the production costs - tend to be higher there.*" Thus, the MR could exhibits higher wages and rents, but this differential would be higher than in the case of the region with natural resource due to the lock-in effect[7]. According to the elements described above, the wages differentials are a function of amenities and rent prices. Following the econometric specification proposed by Roback (1982) and Dumond et al. (1999), the SEM model is specified by:

$$\ln w_{ir} = \beta_0 + X_{ir}\beta_1^t + \beta_2 \ln SHPI_r + \mathbf{AME}_r\beta_3^t + \delta_r \tag{3.3}$$

where w_{ir} is the wage of the worker i in the region r, $\mathbf{X}_{ir}$ is a set of variables representing the HCT, $SHPI_r$ is a spatial housing price index for region r, AME_r is a set of amenities and δ_r is a set of $r-1$ fixed effects for each spatial unit. The strategy consists of how the standard deviation of δ_r is reduced when the $SHPI_r$ and AME_r are incorporated in the analysis. This paper, such as noted earlier, does not follow this approach. However, the specification is presented to keep in mind how the SEM has been tested.

3.3.2 New Economic Geography (NEG)

In contrast to the SEM model, the NEG proposes a positive correlation between wages and MA. The NEG is a spatial general equilibrium model characterized by increasing returns and positive transport costs. The clearing conditions produce three simultaneous equations linking regional wages, total demand and price indexes. However, the empirical exercises have been focused on the estimation of the wage equation. Those regions with higher MA will generate a higher level of wages, giving rise to a core-periphery wage structure.

Formally, it assumes that the consumers obtain utility through the consumption of an homogenous good A and a differentiated good M. The utility function is assumed to be Cobb-Douglas, but the manufactured goods are consumed in two steps. In the first step, the consumer choose varieties of the manufactured goods reflecting a love for diversity. The firms choose the location where they decide to produce and they must pay a transport cost T_{rh} to move the good from r to h. The production function enjoys increasing return to scale; consequently, the firms have incentives to locate the total production in only one plant. Equivalently, the firms prefer to be located close to the large markets. The demand side offers units of labor and the firms

[7] Next sections provide empirical evidence to support this argument.

are willing to use them in the production function. The equilibrium wage is represented as a function of the market access.

$$w_r = \left[\sum_h Y_h I_h^{\epsilon - 1} T_{rh}^{1-\epsilon} \right]^{1/\epsilon} \tag{3.4}$$

where w_r is the wage in the region r, Y_h is the total demand in the region h, I_h is the price index in the location h and ϵ is the preference for diversity. Given this wage equation, the NEG predicts that the wage differential can be explained by the size of the potential market. Following Fingleton (2006, 2008, 2009) and incorporating the worker heterogeneity, the reduced form is established by:

$$\ln w_{ir} = \gamma_0 + \gamma_1 X_{ir} + \gamma_2 \ln \sum_h \frac{Y_r}{T_{rh}} = \gamma_0 + \gamma_1 X_{ir} + \gamma_2 \ln MA_r \tag{3.5}$$

where Y_r represents the income level for spatial unit, T_{rh} is the transport cost between the unit r and h and MA_r is the Market Access of the unit r. The strategy, in contrast to the NEG model, consists in evaluating if the parameter γ_2 is positive and significant. Most estimations for developed countries were in the range from 0.12 to 0.30. If this parameter is significant after controlling by worker heterogeneity, then the MA is assumed to provide the main explanation behind the spatial wage inequality.

3.4 Econometric strategy and data

3.4.1 Econometric strategy

This paper evaluates how the HCT, NEG and SEM compete in explaining the wage differential. Recall that the consideration of these theories generates a scale problem where HCT is presented at micro data level while the NEG and SEM are forces working at a more aggregate level. Simultaneously, most literature has ignored the interaction between micro (worker characteristics) and macro level (MA, amenities and rent prices). This paper overcomes these problems proposing a Multilevel Analysis. The initial point is a hedonic wage equation capturing the variation at Level 1 and 2:

$$\begin{aligned} \ln w_{ir} &= \beta_0 + \varepsilon_{ir} \\ \beta_0 &= \gamma_{00} + U_{0r} \end{aligned} \tag{3.6}$$

where ε_{ir} and U_{0r} present zero mean and variances $var(\varepsilon_{ir}) = \sigma^2$ and $var(U_{0r}) = \tau_0^2$ for the community

r. The term τ_0^2 represents how much is the wage heterogeneity among the r spatial units. Note how equation articulates the spatial wage inequality: each community presents a different wage average with an effect U_{0r} above or below the national mean. The variances τ_0^2 and σ^2 can be estimated from the data using Maximum Likelihood (ML) or Restricted Maximum Likelihood (REML). Both parameters indicate how the variables at Level 1 (HCT) and at level 2 (SEM and NEG) contributes to the explanation of the spatial wage differentials. The correlation between the variance attributed to the Level 2 is captured by the intra-class coefficients:

$$\rho_1 = \frac{\tau_0^2}{\sigma^2 + \tau_0^2} \tag{3.7}$$

If this parameter is close to zero, then there is not information on the Level 2 to be attributed to the variance of the dependent variable and there is no evidence to support the hypothesis of spatial wage inequality. A natural step after the base model consists of estimating the same model, but including the HCT proxies:

$$\ln w_{ir} = \gamma_{00} + \beta_1 educ_{ir} + \beta_2 exp_{ir} + \beta_3 exp_{ir}^2 + \beta_4 educ_{ir} * exp_{ir} + \mathbf{HC}_{ir}\beta_5^t + U_{0r} + \varepsilon_{ir} \tag{3.8}$$

where w_{ir} is the wage per hour, $educ_{ir}$ are the years of schooling, exp_{ir} are the years of experience and HC_{ir} is a vector of worker characteristics such as marital status, economic sector, sex among others. This specification captures the proposition of the HCT and spatial sorting, and the wage inequality is estimated with the new estimate of τ_0^2. Once the variance parameters are estimated, the intra-class effect is computed. This value is an indicator to evaluate if the aggregated spatial scales, where NEG and SEM are operating, have important contributions in the analysis of the wage inequality. If the intra-class coefficient is significant, then a natural question is how much of this inter-group wage variability can be attributed to each theoretical model. This paper uses this argument to evaluate the differences between both models through the specification of the intercept:

$$\beta_0 = \gamma_{00} + U_{0r} + f(NEG) \tag{3.9}$$

$$\beta_0 = \gamma_{00} + U_{0r} + f(SEM) \tag{3.10}$$

This specification implies that the intercepts are different across spatial units and this difference could be explained by the NEG or SEM. For the case of the SEM theory, $f(SEM)$ contains several variables used by the literature such as the housing price index, consumer and producer amenities. The sources described by $f(NEG)$ use two specifications of MA which are described in the next section. This strategy enables evaluation of the research questions number one and two, but this does not permit testing the hypothesis of a combination of both models for different geographical parts of the country. To carry out this test, this paper splits the Chilean territory into five areas known as far north, near north, center, south and extreme south. Each one of these sources is interacted with the sources of NEG to evaluate if one of these theories is relevant for some geographical areas, but not for another. Putting in order the models, the identification strategy is defined as shown in table (3.1).

Table 3.1: Econometric Strategy

	Specification
Model 1	$\ln w_{ir} = \gamma_{00} + U_{0r} + \varepsilon_{ir}$
Model 2	$\ln w_{ir} = \gamma_{00} + U_{0r} + \beta_1 educ_{ir} + \beta_2 exp_{ir} + \beta_3 exp_{ir} + \beta_4 educ_{ir} * exp_{ir} + \mathbf{HC}_{ir}\beta_5^t + \varepsilon_{ir}$
Model 3	$\ln w_{ir} = \gamma_{00} + U_{0r} + \beta_1 educ_{ir} + \beta_2 exp_{ir} + \beta_3 exp_{ir} + \beta_4 educ_{ir} * exp_{ir} + \mathbf{HC}_{ir}\beta_5^t + \varphi_1 e\bar{d}uc_r + \upsilon_1 sex_{ir} + \varepsilon_{ir}$
Model 4	$\ln w_{ir} = \beta_0 + \beta_1 educ_{ir} + \beta_2 exp_{ir} + \beta_3 exp_{ir} + \beta_4 educ_{ir} * exp_{ir} + \mathbf{HC}_{ir}\beta_5^t + \varphi_1 e\bar{d}uc\varepsilon_{ir}$ $\beta_0 = \gamma_{00} + U_{0r}$ $\beta_1 = \gamma_{10} + U_{1r}$
Model 5	$\ln w_{ir} = \beta_0 + \beta_1 educ_{ir} + \beta_2 exp_{ir} + \beta_3 exp_{ir} + \beta_4 educ_{ir} * exp_{ir} + \mathbf{HC}_{ir}\beta_5^t + \varphi_1 e\bar{d}uc\varepsilon_{ir}$ $\beta_0 = \gamma_{00} + U_{0r}$ $\varphi_1 = \gamma_{20} + U_{2r}$
Model 6-7	$\ln w_{ir} = \beta_0 + \varepsilon_{ir}$ $\beta_0 = \gamma_{00} + U_{0r} + \beta_6 MA_r$
Model 8	$\ln w_{ir} = \beta_0 + \beta_1 educ_{ir} + \beta_2 exp_{ir} + \beta_3 exp_{ir} + \beta_4 educ_{ir} * exp_{ir} + \mathbf{HC}_{ir}\beta_5^t + \varepsilon_{ir}$ $\beta_0 = \gamma_{00} + U_{0r} + \beta_6 MA_r$
Model 9	$\ln w_{ir} = \beta_0 + \beta_1 educ_{ir} + \beta_2 exp_{ir} + \beta_3 exp_{ir} + \beta_4 educ_{ir} * exp_{ir} + \mathbf{HC}_{ir}\beta_5^t + \varphi_1 e\bar{d}uc\varepsilon_{ir}$ $\beta_0 = \gamma_{00} + U_{0r} + \beta_6 MA_r$
Model 10	$\ln w_{ir} = \beta_0 + \beta_1 educ_{ir} + \beta_2 exp_{ir} + \beta_3 exp_{ir} + \beta_4 educ_{ir} * exp_{ir} + \mathbf{HC}_{ir}\beta_5^t + \varphi_1 e\bar{d}uc\varepsilon_{ir}$ $\beta_0 = \gamma_{00} + U_{0r} + \beta_6 MA_r$ $\varphi_1 = \gamma_{20} + U_{2r}$
Model 11	$\ln w_{ir} = \beta_0 + \varepsilon_{ir}$ $\beta_0 = \gamma_{00} + U_{0r} + \beta_7 HP_r$
Model 12	$\ln w_{ir} = \beta_0 + \beta_1 educ_{ir} + \beta_2 exp_{ir} + \beta_3 exp_{ir} + \beta_4 educ_{ir} * exp_{ir} + \mathbf{HC}_{ir}\beta_5^t + \varepsilon_{ir}$ $\beta_0 = \gamma_{00} + U_{0r} + \beta_7 HP_r$
Model 13	$\ln w_{ir} = \beta_0 + \beta_1 educ_{ir} + \beta_2 exp_{ir} + \beta_3 exp_{ir} + \beta_4 educ_{ir} * exp_{ir} + \mathbf{HC}_{ir}\beta_5^t + \varphi_1 e\bar{d}uc + \varepsilon_{ir}$ $\beta_0 = \gamma_{00} + U_{0r} + \beta_7 HP_r$
Model 14	$\ln w_{ir} = \beta_0 + \beta_1 educ_{ir} + \beta_2 exp_{ir} + \beta_3 exp_{ir} + \beta_4 educ_{ir} * exp_{ir} + \mathbf{HC}_{ir}\beta_5^t + \varphi_1 e\bar{d}uc + \varepsilon_{ir}$ $\beta_0 = \gamma_{00} + U_{0r} + \beta_7 HP_r$ $\varphi_1 = \gamma_{20} + U_{2r}$
Model 15	$\ln w_{ir} = \beta_0 + \beta_1 educ_{ir} + \beta_2 exp_{ir} + \beta_3 exp_{ir} + \beta_4 educ_{ir} * exp_{ir} + \mathbf{HC}_{ir}\beta_5^t + \varphi_1 e\bar{d}uc + \omega_1 \mathbf{AMEN}_r + \varepsilon_{ir}$ $\beta_0 = \gamma_{00} + U_{0r} + \beta_6 MA_r$ $\varphi_1 = \gamma_{20} + U_{2r}$

3.5 Data

This paper uses information at two levels: the first one provides micro-level or individual worker level information while the second set is at the community level. The micro data information is obtained from the Socio-Economic Characterization National Survey 2009 (CASEN 2009) which is built at the individual level[8]. The selected workers (78,468 in the sample) are between 14 and 64 years and they received a positive hourly wage. The hourly wage is computed by dividing the monthly wage by the number of monthly hours worked. In order to avoid the effects of outliers, the sample selected is bounded by the 1 and 99 percentile of the total sample. Table 3.2 shows the name, a brief description, the metrics and some descriptive statistics for variables at Level 1. Most of these variables are continuous or dummy variables, but the variables related with economic sector, occupation and labor situation are nominal. For these cases, the variables were decomposed in $n-1$ dummy variables, where n being the total number of the categories.

Table 3.2: Descriptive statistics for level 1 and level 2

Variable	Description	Metric	Obs	Mean	Std. Dev.	Min	Max
Wage	Wage per hour	Continuous	78468	7298	11721	22	1211860
Education	Years of Education	Continuous	78468	10.3	3.8	0	20
Age	Age	Continuous	78468	39.6	12.0	15	63
Sex	Sex	1=Men 0= Woman	78468	0.65	0.00	0	1
Head	Household head	1=Household head 0=Otherwise	78468	0.487	0.002	0	1
Married	Marital Status	1=Married 0=Otherwise	78468	0.433	0.002	0	1
Permanent	Permanent Job	1=Permanent 0=Otherwise	78468	0.672	0.002	0	1
Sector	Economic sector such as mining, agricultural among others	9 categories	78468	5.148	3.071	1	9
Occupation	Occupation such as manager, scientific among others	10 categories	78468	7.487	2.402	1	10
Labor situation	Situation such as entrepreneur, employee among others	9 categories	78468	4.279	1.446	1	9
SHPI	Spatial Housing Price index at county level	SHPI	331	1.22	0.38	0.60	3.79
MA wages	Market access with total wages per county	Sum of wages	331	174.35	207.40	14.34	940.42
MA pop	Market access with total population per county	Sum of population	331	0.14	0.20	0.008	0.896
Location 1	Location quotient agricultural and fishing sector	Continuous between 0 and infinite	331	1.12	0.74	0	2.775
Location 2	Location quotient mining	Continuous between 0 and infinite	331	1.02	2.32	0	16.836
Temperature	Standard deviation of temperature	Continuous	331	6.76	1.01	4.274	8.159
Sea	Sea access	1=Sea access 0=Otherwise	331	0.30	0.58	0	7
UV	Number of hours with UV radiation above 5.	Continuous	331	5.57	1.05	2.4	10.6

The information at the county level was obtained from several sources. Some variables are estimated from CASEN 2009 aggregated at to the community level, or from the National System of County Information (SINIM)[9] such as the MA, SHPI and Location Quotients. Most of the amenities variables were obtained from different public data sources and merged in a unique database.

3.6 Results

The results are presented in the table 3.3, table 3.4, table 3.5 and table 3.6. Table 3.3 shows the estimation of the base line: a regression between wages and a random intercept that varies across the geographical units, namely the 331 communities. The key parameter is the intercept variance parameter τ_0^2 that represents how much variability exists around the national average wage. This variance is interpreted as a measure of spatial wage inequality. Table 3.3, table 3.4, table 3.5 incorporate several controls variables derived from the HCT, NEG and SEM respectively, but note that they operate at different scales: HCT at Level 1 and NEG and SEM at Level 2. The Multilevel Analysis determines how the parameter τ_0^2 is affected with

[8]CASEN is a national survey with social and economic information at individual level.

[9]http://www.sinim.gov.cl/

each approach, but with the advantage of considering the interaction between worker characteristics and explanatory variables at the aggregate level. If the approach explains the spatial wage differential, then this will appear as a reduction in the intercept variance. Each approach is compared against the τ_0^2 estimated in the base line, in other words, without any type of explanatory variables. Simultaneously, the approaches compete in the reduction of the residual variance σ^2. Finally, table 3.6 tests whether the NEG explains the wage inequalities for different geographical areas of Chile.

3.6.1 Human Capital Theory and Spatial Sorting of Skills

The first step consists of computing how much wage variability can be attributed to the communities. This variance is estimated through Specification 1 described in table 3.1, and the estimations of Model 1 are presented in table 3.3. The dependent variable is the logarithm of the hourly wage and the explanatory variable is a national intercept (γ_{00}) plus two random errors terms $U_{0r} + \sigma_{ir}$. The wage inequality at the community level is represented by $var(U_{0r}) = \tau_0^2$ and it is significant (at the 1% level) with a value of 0.052, supporting the existence of spatial wage inequality at the community level: 12% of the total wage variability can be attributed to the community level. Model 1 suggests the existence of spatial wage inequality and it constitutes the base line to provide comparisons with the different approaches. This specification is the initial point of the estimation exercise and then this model reduces to 0% the wage inequality.

As described previously, the main source for explaining the spatial wage differential can also be attributed to the Human Capital Theory. CASEN 2009 provides a rich dataset for controlling by observable worker characteristics such as education, age, sex among others variables (see Level 1 in table 3.2). A natural step would be to evaluate how the wage inequality reacts when these variables are added to the model, as, for example, in Specification 2 (see table 3.1). Model 2 (table 3.3) shows the coefficients associated with the proxies of HCT while the categorical variables related with economic sector, occupation and labor situation were left out for space reasons[10]. This column shows the expected signs: the workers with higher education and experience present a positive wage premium and the gender discrimination is represented by a pseudo-elasticity of 21%. The married, household head and workers with permanent job also show higher wages. However, the most interesting information is represented by the estimated variances. First, the spatial wage inequality is reduced by 53.85% (from 0.052 to 0.024). This result supports the hypothesis of spatial sorting: the skill of the workers is not randomly distributed across the space. When the controls are imposed on the worker characteristics, the spatial wage inequality among communities is substantially reduced. This result does not differ from the literature that has identified that spatial sorting explains about 50% of the wage

[10] The complete dataset is available upon author request, Stata code and complete tables are available upon author request.

Table 3.3: Wage equation Human Capital

Dependent variable	Model 1	Model 2	Model 3	Model 4	Model 5
		Fixed Effects			
Constant (γ_{00})	12.291	12.067	11.358	11.477	11.389
S.E	0.013	0.052	0.077	0.097	0.077
Education (β_1)		0.018	0.018	0.018	0.018
S.E		0.002	0.002	0.002	0.002
Age (β_2)		0.014	0.014	0.014	0.014
S.E		0.001	0.001	0.001	0.001
$Age * Age$ (β_3)		-0.000	0.000	0.000	0.000
S.E		0.000	0.000	0.000	0.000
$Age * Education$ (β_4)		0.000	0.000	0.000	0.000
S.E		0.000	0.000	0.000	0.000
Sex (β_5^t)		0.207	0.207	0.206	0.206
S.E		0.004	0.004	0.004	0.006
Household head (β_5^t)		0.084	0.084	0.084	0.085
S.E		0.004	0.004	0.004	0.004
Married (β_5^t)		0.062	0.062	0.062	0.062
S.E		0.003	0.004	0.004	0.004
Permanent Job (β_5^t)		0.201	0.200	0.201	0.201
S.E		0.004	0.004	0.004	0.004
		Random Effects			
Intercept Variance (τ_0^2)	0.052	0.024	0.016	0.029	0.015
S.E	0.004	0.002	0.001	0.003	0.001
Education Variance				0.000	
S.E				0.000	
Sex Variance					0.004
S.E					0.001
Residual Variance (σ^2)	0.368	0.226	0.226	0.224	0.225
S.E	0.002	0.001	0.001	0.001	0.001
Intra-class correlation	0.124	0.096	0.065	0.114	0.078
Log-restricted likelihood	-72650.9	-53654.6	-53592.8	-53527.0	-53555.4
Reduction wage inequality	0.00%	-53.85%	-69.78%	-44.69%	-70.52%
Reduction Residual Variance	0.00%	-38.59%	-38.58%	-39.06%	-38.80%

variation (Combes et al. 2008). Simultaneously, the HCT has a significant impact on the residual variance which is also reduced by 38.59% (from 0.368 to 0.226). With the inclusion of these variables, the intra-class correlation falls from 12.38% to 9.6%; this finding implies that the wage inequality is still related with the communities. Summarizing, the HCT is crucial to evaluate the spatial wage differential: this theory explains almost a 54% of the spatial wage differential.

Alternatively, the HCT affects the wages not only at the individual level; an average effect can also generate a positive impact on productivity. This idea has also been highlighted by the literature of economics of agglomeration through the knowledge spillovers (Marshall, 1920). The Specification 3 (table 3.1), captures this effect with the incorporation of the average education of the community and the results are shown in Model 3 (table 3.3). The hypothesis establishes that those communities with high average education generate more interaction among economic agents and a positive wage impact should be estimated. This impact φ_1 is estimated in 0.071 and that is significant with a p-value lower than 1% and it represents almost six times the impact per year of education. This parameter suggests that wage inequality can be a cumulative process: those regions with better education generate a positive impact on wages that, simultaneously, attract high skill workers and thus may even increase the wage inequality. This cumulative causation has been widely discussed in regional science as one of the main ideas to support the core-periphery structure observed in the spatial distribution of the economic activity (Myrdal, 1957). The inclusion of the average education reduces the spatial wage inequality by 69.78%, but its impact on the residual variance is marginal. With this evidence, the rest of the models always incorporate the role of this aggregated variable.

In addition to differences in the random intercept, the slopes can also vary across space. For example, knowledge spillovers may be more pronounced in dense geographical areas. Considering this argument, large cities should generate higher impacts on productivity. This implies that the returns to education could also vary according to different city size. The Specification 4 (table 3.1), captures this idea and its estimation is shown in Model 4 (table 3.3). The $var(U_{1j})$ is close to zero (0.0000861) which does not provide statistical evidence to support a differential in the returns to education by community. A potential explanation for this result is the rich sets of controls associated with education that are used in the estimation. Thus, most of the heterogeneity in education is captured by the fixed effects.

Finally, a potential source of wage variation per community is based on the role of the gender gap in wages. For example, the geographical units located in the north side of the country (see figure 3.2) are strongly related with the mining industry, an activity where the female participation is low. For this reason, gender could present a spatially random effect such as is the one shown in Specification 5 (table 3.1) and tested in Model 5 (table 3.3). The incorporation of the random slope for sex reduces the wage inequality by

70.52% (from 0.052 to 0.015) and the sex variability reduces the intra-class coefficient up to 0.078 that means that only a 7.8% of the total wage variability is attributed at community level. This finding supports the hypothesis that the gender gap plays a role in explaining the spatial wage inequality that could be strongly related with the spatial location of industries grounded in natural resources production. Summarizing the findings of table 3.3, the spatial sorting and the gender gap play a fundamental role for explaining the spatial wage inequality. Both sources explain almost 71% of the wage variation. From the suspected variance attributed to the level 2, only the 29% remains unexplained after controlling for worker characteristics and the gender gap. Now, the rest of the paper proceeds to explain the remaining 29% of wage inequality by using the two described theories, namely NEG and SEM.

3.6.2 New Economic Geography (NEG)

According to table 3.3 focusing on the Human Capital Theory, average education and the gender gap are relevant for explaining the wage differential and they will now be a control to evaluate the NEG impact. However, in order to analyze the role of the NEG by itself, Model 6 and 7 (table 3.4) only include the MA as explanatory variables. Models 6 and 7 contain two proxies for the MA. The MA is estimated as follows:

$$MA_r = \sum_h Y_r/d_{rh} \tag{3.11}$$

Such as most of the literature, both proxies use the distance as a proxy of the transport cost between the community r and h. However, this paper considers the limitations of the geographical distance, especially in the south of Chile where the presence of islands suggest that the geographical distance may not represent the transport cost properly. For this reason this paper uses the driving distance instead of the geographical distance. Only a small set with missing values was imputed with geographical distances between centroids. The difference between both models is with respect to the variable Y_h, where Models 6 and 7 use the total population and total income by community, respectively. The correlation between both variables is extremely high (around 0.96) and they are also represented in figure 3.2. High values of MA are extremely concentrated around the Metropolitan Region of Santiago. Both equations are represented by Specification 6, (table 3.1) and the estimations are shown in Models 6 and 7 in table 3.4. The estimations suggest a wage-MA elasticity of 6% for both specifications. This coefficient is extremely low in comparison to those estimated for most developed countries where values between 12% and 30% were found (Garcia, 2006). The results here support the hypothesis about the discrete role played by the MA in the Chilean case. This theory only helps to reduce the spatial wage inequality by approximately 6% (from 0.052 to 0.049) for both specifications. Given that MA built with population shows a slight better performance than total income (see Log-likelihood), the rest of the estimations are computed with the population proxy. From this perspective, the wage-MA elasticity is extremely low and also the reduction in wage inequality, which seems to support the hypothesis that the MA does not provide significant explanation in the Chilean case.

Table 3.4: Wage equation for NEG

Dependent variable	Model 6	Model 7	Model 8	Model 9	Model 10
		Fixed Effects			
$MA\ (\beta_6)$	0.059	0.062	0.043	0.004	0.008
$S.E$	0.012	0.014	0.008	0.008	0.008
		Random Effects			
$Intercept\ Variance\ (\tau_0^2)$	0.049	0.049	0.022	0.016	0.015
$S.E$	0.004	0.004	0.002	0.001	0.001
$Education\ Variance$				0.000	
$S.E$				0.000	
$Sex\ Variance$					0.004
$S.E$					0.001
$ResidualVariance\ (\sigma^2)$	0.368	0.368	0.226	0.226	0.225
$S.E$	0.002	0.002	0.001	0.001	0.001
Intra-class correlation	0.117	0.118	0.088	0.065	0.078
Log-restricted likelihood	-72642.9	-72644.5	-53644.9	-53596.7	-53558.8
Reduction wage inequality	-6.18%	-5.11%	-58.09%	-69.72%	-70.56%
$Reduction\ Residual\ Variance$	-0.12%	-0.12%	-38.58%	-38.58%	-38.80%

In order to be consistent with table 3.3, the HCT, average education and gender gap must be considered to evaluate the NEG. The Specification 8 (table 3.1) and the estimations for Model 8 (table 3.4) incorporate the HCT. Now the MA impact is even lower, with an elasticity of 4.3% and the reduction of 58% in the wage inequality due almost totally to the HCT. The residual variance is very similar to the residual variance estimate for Model 5 reinforcing the finding that NEG does not help to explain the spatial wage inequality. Finally, the average education and the sex random effects must be added in Specifications 9 and 10, (table 3.1). Model 9 (table 3.4) shows that the inclusion of average education reduces almost to zero the wage-MA elasticity that is no longer significant now. The wage inequality is reduced by 69.72%, but this effect is mainly attributed to the HCT. For the case of the sex random effect, the situation is very similar and the wage inequality is reduced almost by 71%. Note that this reduction is almost similar to that provided by Model 5. In other words, the NEG does not seem to explain the wage inequality beyond the explanation provided by the HCT.

3.6.3 Spatial Equilibrium Model

After testing the role of the NEG, the SEM is used now to explain the spatial wage variation. The strategy to estimate the role of the SEM is very similar to that used for NEG, but here wages, housing prices and amenities are related. This paper considers that housing price plays a crucial role for the Chilean case, and Models 11, 12, 13 and 14 (table 3.5), only consider the role of housing prices. After estimating the role of housing prices, Model 15 adds a set of consumer and producer amenities.

The first step is the consideration of the SHPI revealed by Specification 11 (table 3.1) and the results are presented in Model 11 (table 3.5). The first result suggests a wage-SHPI elasticity of 0.614, significant at the 99% confidence level. This estimation supports the intuition derived from figure 3.2 where the wages and SHPI are strongly related. This result is interesting for two reasons. First, this positive correlation reveals the existence of spatial price differentials and why this measure is relevant for explaining the spatial inequalities. The communities with higher housing prices are compensated by higher wages, strongly suggesting that the nominal comparison is not appropriate. However, the SHPI is not provided by any statistical office, and there is no official price index for the regions. This lack of information makes it difficult to conduct an appropriate regional analysis. A second consequence derived from this result reinforces the accuracy of the price index estimated by the microeconomic approach presented in Iturra and Paredes (2011). This index demonstrates superior outcomes than other alternative measures due to the extremely high elasticity with wages (0.61)[11]. Regarding wage inequality, SEM shows a much better performance than NEG: the SHPI by itself reduces the wage inequality by 57.69%. This is the first result in support of an argument that the SEM model seems to fit better than NEG for the Chilean case.

The second natural step is the incorporation of the HCT and the average education effect described by Specifications 12 and 13 (table 3.1) and the results are shown in Models 12 and 13 (table 3.5). The role of the housing price is considerably reduced, but it is still significant. The SHPI-wage elasticity is between 20 and 30% when the HCT is incorporated. The reduction of the elasticity can be attributed to the correlation between the worker characteristics and the housing quality; those professionals with high wages have a higher propensity to buy more expensive housing with greater number of attributes. This elasticity is estimated for the first time in the Chilean case and the estimations show the relevance of this index for evaluating the wage inequality. The model with average education explains 73.08% of the wage inequality. With respect to the incorporation of a random intercept in the sex variable, Model 14 (table 3.5) includes the gender gap. The incorporation of this random effect is significant again and these factors together explain 75% of the wage inequality. Considering the results of these models, the SEM, in comparison with the NEG, may

[11]See Iturra and Paredes (2011) for details about the comparison of microeconomic approach with other alternatives.

Table 3.5: Wage equation for SEM

Dependent variable	Model 11	Model 12	Model 13	Model 14	Model 15
		Fixed Effects			
$SHPI\ (\beta_7)$	0.614	0.330	0.210	0.206	0.216
$S.E$	0.030	0.025	0.031	0.031	0.031
$L.Q.Agriculture$					0.024
$S.E$					0.015
$L.Q.Minning$					0.008
$S.E$					0.003
$Standard\ Deviation\ Temperature$					0.045
$S.E$					0.008
$Sea\ Access$					-0.010
$S.E$					0.012
UV					-0.027
$S.E$					0.007
		Random Effects			
$Intercept\ Variance\ \tau_0^2$	0.022	0.015	0.014	0.013	0.011
$S.E$	0.002	0.001	0.001	0.001	0.001
$Sex\ Variance$				0.003	0.003
$S.E$				0.001	0.001
$ResidualVariance\ \sigma^2$	0.368	0.226	0.226	0.225	0.225
$S.E$	0.002	0.001	0.001	0.001	0.001
Intra-class correlation	0.056	0.062	0.058	0.066	0.062
Log-restricted likelihood	-72519.3	-53587.5	-53573.9	-53537.9	-53535.3
Reduction wage inequality	-57.69%	-71.15%	-73.08%	-75.00%	-78.02%
$Reduction\ Residual\ Variance$	0.00%	-38.59%	-38.59%	-38.86%	-38.79%

be seen to be the superior model. The elasticity associated with housing prices is higher and the spatial wage inequality is reduced by a 5% over the NEG model. However, the role of the amenities has not been incorporated yet.

The amenities are specific characteristics of the spatial unit, but they are not subject to manipulation by human intervention. Obviously, the researcher cannot capture the complete set of amenities that can affect the producer and consumer side. Given these constraints, the literature prefers to analyze whether the amenities reduce the wage inequality, but it is impossible to account for the complete role of the amenities. A particular type of producer amenity is the existence of natural resources that can produce higher productivity due to economies of location or specialization. Under this scenario, those spatial units with a higher location quotient for the primary sector could present high wages even without evidence of large MA. Model 15 shows how those natural resources intensive communities also present a positive impact on local wages. Both impacts are significant, supporting the hypothesis about the role of the natural resources in developing countries. This coefficient suggests that MA is not sufficient in explaining wage differentials, when natural resources are a crucial type of amenity such as in the case for many developing countries.

A set of consumer amenities was also incorporated, such as the standard deviation of temperature, sea access and the number of hours with UV radiation (see table 3.2). The sea access is not significant for the Chilean case and the UV radiation shows a strange sign: those communities with high UV present lower wages. This strange result can be attributed to the proxy because only some communities present measurement centers of UV radiation. In order to fill the missing data, the UV radiation was imputed to those closest communities with the measurement center. In sum, SEM reduces the wage inequality reaching up to 78.02%. As previously discussed, the main source for explaining the spatial inequality is provided by the HCT and gender gap with (accounting for 70%). The NEG has a negligible explanation power and the SEM approach increases the explanation until up to 78%.

Clearly, the SEM shows a better performance than NEG in contributing to the explanation of spatial wage inequalities in the Chilean case. The wage-MA elasticity is extremely low in Chile, below the average of developed countries. This result supports the hypothesis than there are other alternative sources, particularly natural resources, which could be playing a crucial role in the case of developing countries. While the MA is important in some cases, natural resources dependent countries seem to be a different case. In Chile, the spatial location of natural resources does not coincide with dense areas, and, consequently, higher wages could be differently distributed across the space. Moreover, the particular geographical shape does not contribute to an argument that factor mobility reduces the regional differentials. Thus, regional labor markets could present the case where several theories are working simultaneously across the geographical space. Hence, the final research question addresses an hypothesis than although the MA does not explain the whole Chilean case, this theory could be helpful in the portion of the country where the population is highly concentrated. Since, the center and south center of the country exhibit the highest levels of population density, the MA could improve its performance if the wage-elasticity is estimated differently for these areas. This exercise is estimated and presented in table 3.6.

Table 3.6: Wage-MA elasticity for different geographical areas

Dependent variable	Model 16	
	Coefficient	Population
I	0.072	300301
II	0.018	561604
III	0.067	276480
IV	0.132	698018
V	0.170	1720588
RM	0.121	6745651
VI	0.142	866249
VII	0.162	991542
VIII	0.175	2009549
IX	0.163	953835
X	0.097	815395
XI	0.016	102632
XIV	0.133	376704
XV	0.093	187348
Random Effects		
$Intercept\ Variance\ (\tau_0^2)$	0.020	
$S.E$	0.002	
$Sex\ Variance$	0.060	
$S.E$	0.005	
$ResidualVariance\ (\sigma^2)$	0.353	
$S.E$	0.002	
Intra-class correlation	0.186	
Log-restricted likelihood	-71327.834	
Reduction wage inequality	61.22%	
$Reduction\ Residual\ Variance$	4.14%	

The results in table 3.6 are the outcomes from the test of the hypothesis derived from the third research question: 3) can Chile be a case where the MA reveals a different degree of explanation across the country? The table shows interesting patterns. First, the wage-MA elasticity is different across the space where the range of the estimated coefficients is between 1.6% and 17.5%. The lower parts of this range is associated with the geographical location of the regions that are copper exporters, particularly the XV, I, II and III regions, with an elasticity smaller than 10% (see figure 3.2). For these areas, the MA does not seem to contribute substantially to explaining the wage differentials such as it was established in the initial statements. A different situation is observed between the region IV and XIV where the wage-MA ranges between 12% and 17.5%. This result seems to support the hypothesis that MA is a relevant explanation only in those regions located around the MR, such as the center-south, which accommodates 40.62% of the national population. Only for this set of regions does the MA presents a significant relationship and these estimations argue that the NEG provides a causal mechanism only for this subset of regions. However, even here, the elasticity is extremely low with respect to the developed countries. Figure 3.5 shows clearly the correlation between the size of the region (measured by population) and the wage-MA elasticity. Those regions characterized by an economic productive system based on natural resources are ones where the elasticity is particularly lower than 10%. Summarizing, by itself, the NEG is not sufficient to account for the whole Chilean case. The national wage-MA elasticity is lower than the same estimates for developed countries, but this theory improves its performance when the regional heterogeneity is included. In the last case, the NEG improve its fit in the dense regions, but it is not enough for those regions with natural resources where the fit is extremely low.

3.7 Conclusions

Spatial wage inequality in LA countries can best be understood and empirically estimated by recognizing its particular context. Key to this context is the role played by natural resources. In Chile, the spatial wage inequality has not been widely discussed even though the data indicate one of the highest levels of spatial inequality in the world. In the investigation of inequality, this paper asks the question whether directly importing economic models from developed countries is appropriate for the case of countries with such high dependence on natural resources. Based on natural resource dependency and the particular physical geography of Chile, this paper compares the adequacy of NEG versus SEM models of spatial wage differentials, in both cases controlling for HC. In order to capture simultaneously the underlying micro contributions of HCT and the influence of spatial characteristics, this paper estimates spatial wage disparities as a function of both, the latter represented first by NEG and then by SEM.

A summary of the performance for each theory is detailed in the figure 3.5. Twelve percent of the total wage variability can be attributed to the community level, but it is not independent of HCT. The HCT is incorporated with several proxies such as education, experience, sex, among others, and all of these proxies show the expected signs. However, the most interesting information is represented by the estimated variances. First, the spatial wage inequality is reduced by 54% (from 0.052 to 0.024) when the HCT is

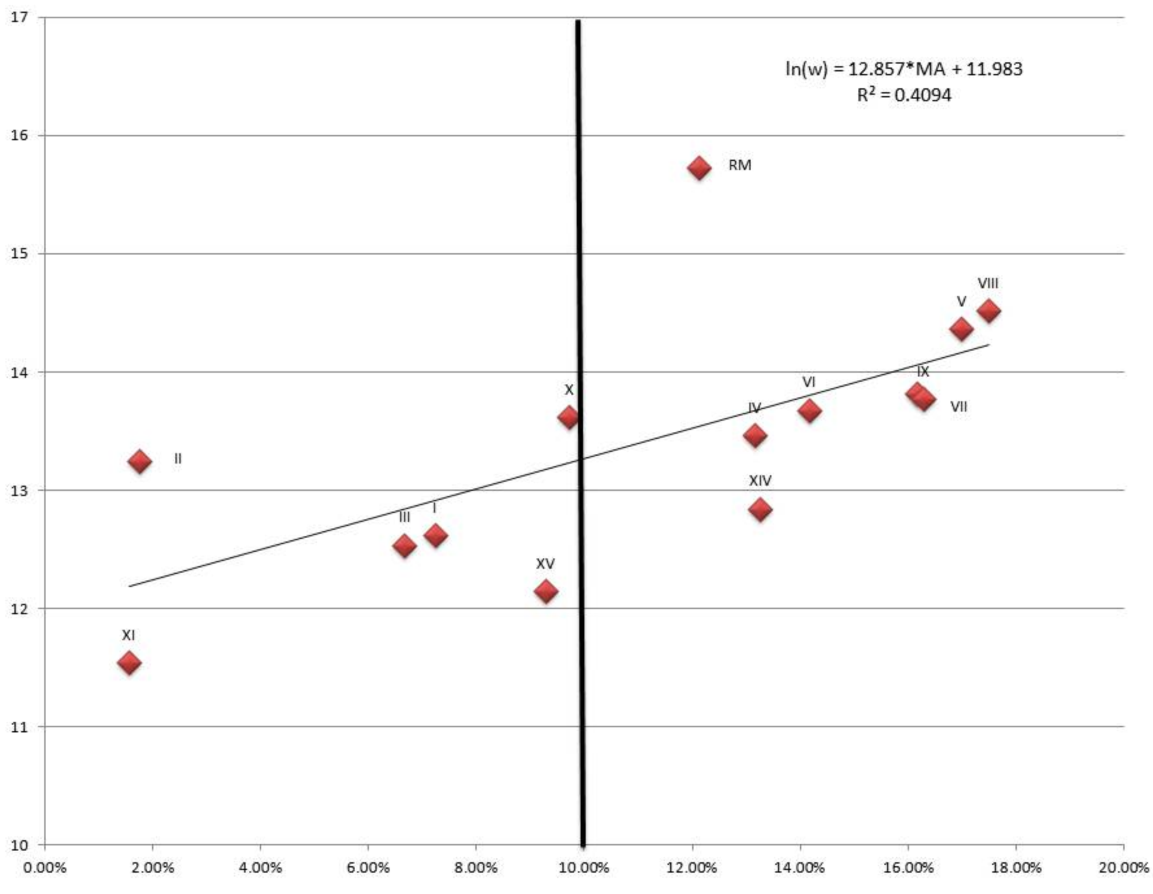

Figure 3.5: Wage-MA elasticity across Chilean regions.

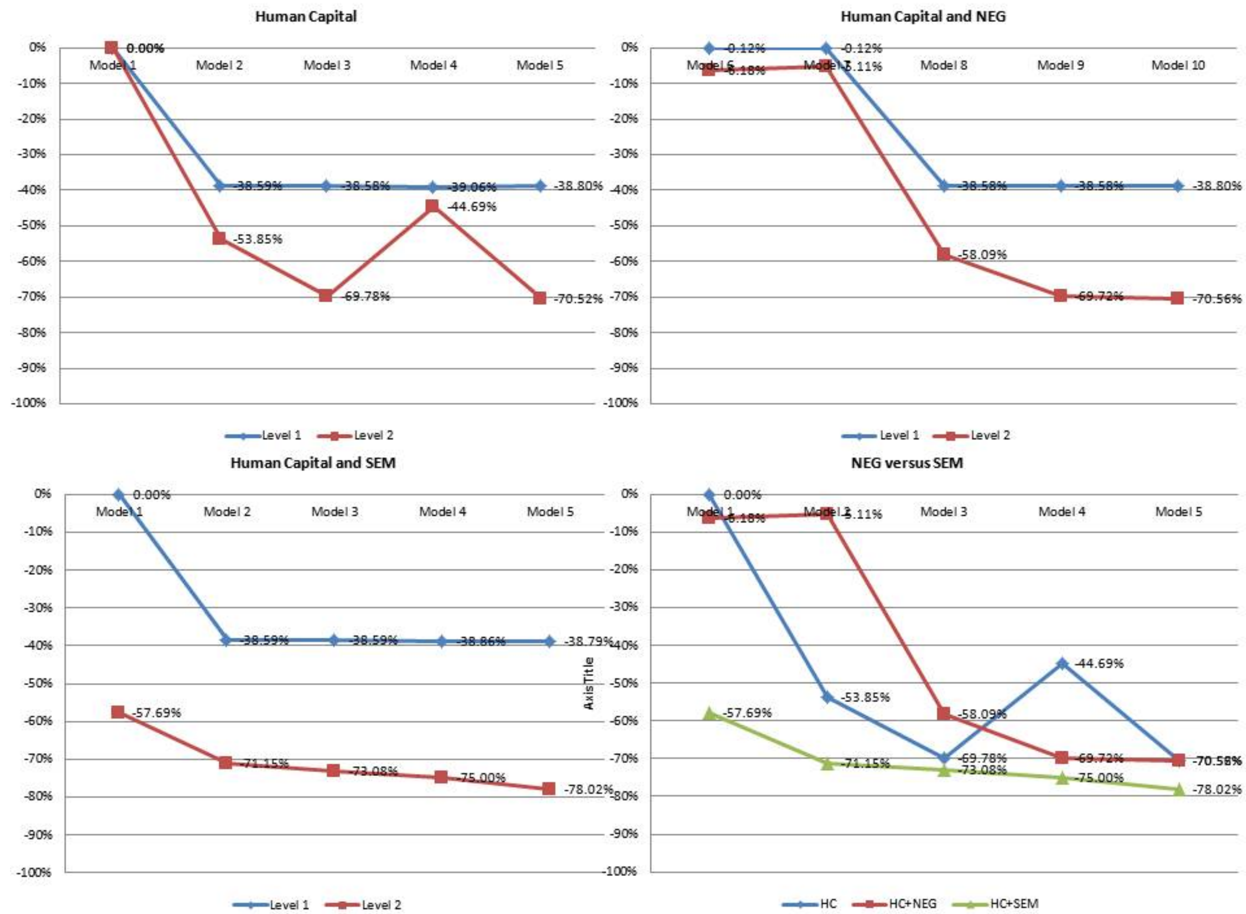

Figure 3.6: Spatial wage and residual variance reduction.

included. This result supports the hypothesis of spatial sorting: the skill of the workers is not randomly distributed across the space. When the controls are imposed on the worker characteristics, the spatial wage inequality among communities is substantially reduced. The inclusion of the average education reduces the spatial wage inequality by 69.78% and the random slope for sex reduces the wage inequality by 71% (from 0.052 to 0.015). Summarizing, the HCT reduces the variability attributed at community level and it explains almost a 71% of the spatial wage variability with only 29% unaccounted for at the community level. The estimations suggest a wage-MA elasticity of 6% for both specifications (namely, population and total income) specifications and this coefficient is extremely low in comparison with the most of developed countries. When the HCT is incorporated, the wage-MA impact is even lower, with an elasticity of 4.3% and the reduction of 50% in the wage inequality. The consideration of average education and random effect for sex reduces it to a total of 71%. According to these results, the NEG does not seem to help to explain the wage inequality beyond the explanation provided by the HCT.

The performance played by SEM is significantly better than NEG. Results suggest extremely high wage-SHPI elasticity with this coefficient as 0.614. With respect to the wage inequality, SEM shows a much better performance than NEG: the SHPI by itself reduces the wage inequality by 57.69%. The model with average education explains 73.08%. Adding a random intercept increases the values to 75% while the inclusion of consumer and producer amenities enables a total of 78.02% of the wage inequality to be explained.

According to the estimation, the NEG could improve its performance if a spatially differentiated effect is allowed. First, the wage-MA elasticity is different across the space with a range of the estimated coefficients between 1.6% and 17.5%. This range is associated with the geographical location of the country where the regions that are copper exporters, particularly the XV, I, II and III regions, show elasticity lower than a 10%. For these areas, the MA does not seem to be sufficient for explaining the wage differentials such as it was established on the initial statements. A different situation is observed between the regions IV and XIV where the wage-MA ranges between 12% and 17.5%. This result seems to support the hypothesis that MA is a relevant explanation only in those regions located around the MR where 40.62% of the national population resides.

Summarizing, the case of developing countries characterized by natural resources present evidence to support a different causal mechanism than the provided by NEG. For the Chilean case, the spatial wage inequality is much better explained by amenities than Market Access. The MA role is not sufficient for this case, and its role could be relevant only for the large agglomerations in the country. This paper presents contributions to further understanding the Chilean case, providing evidence to guide the future agenda to analyze the wage differential in those countries where the natural resources assume a significant role in the productive system.

Chapter 4

A methodology to compute

Road map for Chapter 4: The Section 4.1 discusses the relevance of Cost of Living for regional analysis. The relevance of the Cost of Living is demonstrated for the comparison of real spatial economic magnitudes and the correct evaluation of regional policies. Section 4.2 indicates the lack of discussion of the literature about the inter-regional comparison of housing prices. Particular, there is not discussion about how we can compare the price of heterogeneous goods in the different spatial units. The Section 4.3 proposes a novel methodology to compare housing price using matching estimator methods. This method allows to reduce the problems derived from the heterogeneity in housing characteristics across the space. Section 4.4 presents the data Chile for the estimation process. The Sections 4.5 and 4.6 suggest the the matching estimator reduces the spatial heterogeneity. Moreover, the housing price index proposed is significant different than standard mean price comparisons.

4.1 Introduction

The cost of living is known to vary across regions in most countries; these variations have important implications for studies of regional income convergence, as well as for adjusting national wage bargaining agreements that take regional price differences into account. However, in many countries, regional cost of living indices are not estimated on a regular basis. In this paper, some initial explorations with the development of a regional housing index will be explored. A price index would be helpful in the formulation and design of housing policies, social housing programs or any public policy focused on regional housing markets. Therefore, to design adequate regional housing policies and to understand the dynamics of the housing market, regional scientists must be able to precisely estimate the housing price index across different regions or spatial units. The contribution of this paper is to take heterogeneity into account in the comparison of regional dwellings by using a quasi-experimental control group method (Rosenbaum and Rubin, 1983). Using this method, we match dwellings with similar characteristics and quality between different regions. The output will be two samples (one for each region) of houses with homogeneous characteristics; this will allow comparison of house prices through a regional housing price index. This information will be used to explore two issues: (1) whether housing prices are statistically different across the regions; and (2) the role of inherent heterogeneity in housing in computing these differences.

In this paper, we use three kinds of quasi-experimental control group methods: (1) nearest matching on the propensity score, (2) Mahalanobis[1] matching including the propensity score and (3) Mahalanobis matching within score calipers. This paper evaluates the three methods in the context of the housing price index and chooses the best method based on the reduction of the average regional bias,measured through the standardized differences of house characteristics between spatial units (Tritchler, 1995).

The matching method allows identifying one "control house" in the metropolitan region for each "treatment house"in any region[2] both having statistically similar characteristics. Therefore, one generates two samples with the same number of observations. Hedonic regressions are run on these samples, estimating the hedonic coefficients for the characteristics in each matched sample. Using the Spatial Fisher Housing Price Index and its superlative property, a regional housing price index in Chile is calculated for 2006. The results show that Mahalanobis matching within score calipers was the best method to reduce geographic bias for each covariate among regions and the regional differential in the propensity score in the Chilean case. In addition, the Regional Fisher Price Index shows that the second Region (Antofagasta) is the most expensive in the country. This paper contains six sections. Section 2 surveys the relevant regional housing price index literature. Section 3 discusses the quasi-experimental methods used and presents the hedonic functions. Section 4 describes the data used. Section 5 reports the estimation of quasi-experimental methods and parameters for hedonic regression. Section 6 considers the policy implications of the Regional Housing Price Index. Finally, Sect. 7 presents the conclusions.

4.2 Reviewing the regional price index literature

Rosen (1974)'s work provided the basis for the use of hedonic regression as the principal tool to investigate housing prices. Although there is broad consensus about the appropriateness of hedonic regression to build a house price index, most of the applications have not considered geographic heterogeneity in housing. However, recent literature offers some advances in this area. Mills and Simenauer (1996) proposed a regional house price index for the United States using national data and incorporating geographic heterogeneity through regional fixed effects. The paper revealed significant differences between the index published by

[1] This distance measurement, unlike Euclidean distance, does not depend on the scale of variables and considers the correlations among variables.

[2] There are 12 regions in Chile and one Metropolitan Region. Given that the Metropolitan region is the largest, each house in the regions was matched to a "control house" in the Metropolitan Region.

the National Association of Realtors (NAR) and their proposed regional index. Despite efforts to construct a regional index, the authors did not calculate a hedonic regression for each region, thereby assuming homogeneous behavior of the parameters over space. A different approach was taken by Forret (1991), who partially incorporated geographic heterogeneity in his construction of the regional house price index in 1985 for regions in England. The author recognized regional heterogeneity in housing, arguing that regional differences in price may stem from the neighborhood and physical characteristics of respective regional housing stocks, but did not use hedonic regression to solve the econometric problem associated with geographic heterogeneity. The importance of regional heterogeneity can appear at the intra-regional spatial levels, for example, within metropolitan areas. Thibodeau (1989) computed a tenure specific hedonic housing price index for 60 metropolitan statistical areas (MSA) in the US. Calculating the hedonic coefficients for each MSA, houses were "priced" at market value. Finally, the average ratio of the housing value was estimated for each MSA, in time and space, reflecting the housing price index. His paper computed the hedonic regression for each MSA, but the interpretation of the regional index is debatable because the coefficients used to calculate the value of houses came from heterogeneous housing data. In other words, the hedonic coefficients were calculated with regional housing data, but without any attempt to reduce geographic bias. In spite of great strides made in this area, the spatial aspect still requires more consideration in constructing a regional house price index. Costello and Watkins (2002) highlighted the importance of local house price indices. According to them, the index must consider the minimum geographic scale available and respect the differences among urban markets. According to Paredes and Aroca (2008), geographic heterogeneity could be reduced by using the quasi-experimental control group method. Before building the regional price index, the authors calculated the bias between any pair of regions in Chile. The bias is the average difference in the independent variables to construct the price index. Using a nearest neighbor-matching estimator, the authors matched similar houses in two regions. Finally, they calculated a regional Fisher price index using hedonic regression and incorporated a methodology to reduce regional heterogeneity. However, they did not explore whether quasi-experimental methods fit better for this purpose.

In summary, only a few articles focus on building a regional housing price index that considers spatial heterogeneity. Mills and Simenauer (1996) and Thibodeau (1989) marginally incorporated the regional aspect, but did not make the effort to embrace the spatial dimension. Forret (1991) identified the heterogeneity problem and developed a measurement of the bias, but did not provided a solution for it. Paredes and Aroca (2008) demonstrated an alternative method, but did not test it with alternative quasi-experimental control groups. Quasi-experimental control groups may provide the methodology to construct the regional price index; however the present article will propose different quasi-experimental methods to reduce the bias demonstrated by Paredes and Aroca.

4.3 Methodology

4.3.1 Matching estimator methods

Quasi-experimental methods estimate the effect that a treatment would have on a unit that in fact does not receive the treatment (Rosenbaum and Rubin, 1983). Formally, the literature calls the treated group the "treated group" and the group without treatment as the "potential control group"(Rubin, 1976). In the regional housing market, the treatment can considered the presence of a house in a specific region. To know the regional differential price of the house, the ideal situation would be to have exactly the same house in two different regions at the same time, but this comparison is impossible. The alternative is that each regional house is compared to a similar house in the benchmark region. The comparison is realized between each region $i = 1, 2, ..., n$ and the benchmark region mr, taking into consideration that each house belonging to $i = 1, 2, ..., n$ could have different characteristics from the one in the benchmark region. Given that the Metropolitan Region has the highest number of observations, it will always be possible to find a clone for any regional house. Therefore, the Metropolitan Region is similar to the non-treated region. Let x characterize the vector for a specific house, and the z variable indicate whether the house is located in a specific region i ($z = 1$) or in the benchmark region (metropolitan area) ($z = 0$). The propensity score $e(x)$ is defined as a conditional probability to be located in the region i, given the covariates, that is to say, $e(x) = Pr(z = 1|x)$. The matching using $e(x)$ will balance the distributions of x between the region i and the benchmark region (Rosenbaum and Rubin (1983)). At this point, the matching estimator presents two issues. First, the functional form for Pr is unknown, therefore it must be estimated from the available data. Secondly, $e(x)$ has a continuous metric, and as it is impossible for two $e(x)$ to match exactly, it is therefore necessary to choose an objective criterion to match similar $e(x)$. The literature highlights different functional forms for the probability. For the binary case, Smith (1997) established that there are no critical differences for the popular logit and probit densities, excepting for the constraint on the data generation process where the probabilities are independent of irrelevant alternatives imposed by the logit. The present paper works with a probit model to estimate $e(x)$. Regarding the second issue, this paper compares three methods to match the propensity score: nearest-neighbor matching (NNM), Mahalanobis metric matching with propensity score (MMPS) and Mahalanobis metric within propensity score calipers (MMWPS). The adjustment for each matching method is evaluated using standardized differences. Specifically, the method with the greatest bias reduction measured through standardized differences will be chosen as the best-fit model.

4.3.2 Nearest-neighbor matching

The NNM matches each house belonging to a particular region with a house in the benchmark region mr with a similar propensity score $e(x)$. Assume $e_i^k(x_i^k)$ represents the propensity score of house k in region i considering the covariates x_i^k. Let $e_{mr}^n(x_{mr}^n)$ represent the propensity score of house n in the benchmark region[3] (Metropolitan Region), based on covariates x_{mr}^n. N_i and N_{mr} are the number of observations of regions $i = 1, 2, ..., 12$ and the benchmark region mr, respectively. Putting all this together, houses are matched using the equation

$$C^{NNM}(e_i^k) = \min_n \left\| e_i^k - e_{mr}^n \right\|, n \in N_{mr} \tag{4.1}$$

In this case, $\|(.)\|$ is either based on comparing the index function or is obtained through a distance metric. The control is selected on the basis of the smallest difference between propensity scores given by the Euclidean distance between them. This matching selects a control observation just once for each house k region i , therefore control houses are drawn without replacement. In addition, the control groups are built independently for each region, so a house in the benchmark region can belong to more than one control group.

4.3.3 Mahalanobis metric matching including the propensity score as a covariate

The Mahalanobis metric matches two observations using the Mahalanobis distance of the covariates. This method, instead of minimizing the difference in the propensity scores between treated and control observations, finds a control individual for each treated observation with the closest characteristics estimated through the Mahalanobis distance. This method will eliminate the matching between two observations with different propensity scores, even if the distance between these propensities is small. The MMPS uses the Mahalanobis distance, but includes the propensity score as a covariate. Houses are matched using the equation:

$$DM = (x_i^k - x_{mr}^n)^T C^{-1} (x_i^k - x_{mr}^n) \tag{4.2}$$

where x_i^k and x_{rm}^n includes $e_i^k(x_i^k)$ and $e_{mr}^n(x_{mr}^n)$. Rubin and Thomas (2000) recognized the advantages

[3]The benchmark is Santiago de Chile. It is the political capital and has the highest population in the country. This implies more probability to find a “clone”for regional houses.

of including $e(x)$ in NNM, especially in handling possible problems arising from selection bias.

4.3.4 Mahalanobis metric within propensity score calipers

The MMWPS is a hybrid method that defines a subset of control candidates using the propensity score as a caliper, and selects the control using the Mahalanobis metric on the covariates, including the propensity scores (as MMPS). Sequentially, the first step for MMWPS is to find the closest controls using the propensity score according to the caliper defined.With this subset defined, the possibilities of finding a control with the Mahabalobis method are higher than with the other two proposed matching techniques. The second step for MMWPS is similar to MMPS, namely to match houses with similar Mahalanobis metrics using the covariates and propensity score to compute the metric. The key element of this method is the definition of the caliper scalar. Rosembaum and Rubin (1985) suggest choosing the caliper based on the variance of the propensity score for each group. If σ_i^2 i is the variance for the propensity score in each region $i = 1, 2, ..., n$ and σ_{rm}^2 rm is the variance for the benchmark region, the caliper should be a function of $\sigma = [(\sigma_i^2 + \sigma_{rm}^2)/2]^{1/2}$. This paper follows Cochran and Rubin (1973), who recommended a caliper width of $c = 0.2\sigma$.

4.3.5 Hedonic price and spatial fisher price index

The estimation of hedonic prices is the standard methodology for studying heterogeneous goods. Rosen (1974) provided the theoretical background for the interrelationship among supply, the consumer bidding and hedonic prices. In the particular case of Chile, there are 13 regions. Therefore this article estimates 12 hedonic regressions for each region $i = 1, 2, ..., 12$, with the capital (Metropolitan Region) as the benchmark mr. In Eq. (3.3), P_i^k is the price of the house k in region i and x_{ij}^k is one of the J characteristics for house k in region i . The coefficients β_{0i} and β_{ij} vary for each region i. Equation (4) P_{mr}^{nki} mr represents the price for the housing n of the benchmark region mr matched with the housing k in the region i and x_{mr}^{nki} is one of the J characteristics for the house n in mr, matched with the housing k in the region i. The coefficients β_{mr0}^i and β_{mrj}^i are the hedonic prices obtained for the control group of region i. Given that the three matching methods imply that each house k in the region must have one clone in the Metropolitan Region, then after matching $k = n$:

$$\ln P_i^k = \beta_{i0} + \sum_{j=1}^{J} \beta_{ij} x_{ij}^k + \epsilon_i^k \tag{4.3}$$

$$\ln P_{mr}^{nki} = \beta_{mr0}^i + \sum_{j=1}^{J} \beta_{mr}^i x_{mrj}^{nki} + \epsilon_{mr}^{nki} \tag{4.4}$$

The key to regional comparison is the estimate of the different beta coefficients. Applying the matching estimator method, the x_{ij}^k covariates will be statistically similar to x_{mrj}^{nki} for each region $i = 1, 2, ..., 12$, therefore the difference in the coefficients can be attributed to price differences rather than quality differences. In particular, estimating a price index requires comparing costs across regions using the same basket. In this case, the matching estimator method imposes the same average basket a priori to the estimation of hedonic prices. In this case, price differences can be attributed to differences in the cost of living because the same basket is being evaluated. Once the coefficients for each region and the control group have been estimated, the next step is to estimate the value of housing considering the different characteristics. Previous research has shown that the advantage of hedonic regression is that it can maintain quality constants. However, spatial comparison demands attention to the influence of the basket of characteristics on the value of houses in both regions.Matching reduces the differences between regional baskets, but does not completely eliminate them. This problem has been broadly documented in temporal price indexes, such as the Paasche and Laspeyres price indices. This paper proposes a geometric mean of these two indices to estimate the regional housing price index, as follows:

$$\ln(I_{i/mr}) = 0.5(\ln(\bar{P}_{mr}^i) - (\beta_{i0} + \sum_{j=1}^{J} \beta_{ij}\bar{x}_{mr}^i)) + ((\beta_{mr0}^i + \sum_{j=1}^{J} + \sum_{j=1}^{J} \beta_{jmr}^i \bar{x}_i - \ln(\bar{P}_i)) \tag{4.5}$$

where $\bar{P}_{mr}^i$, $\bar{x}_{mr}^i$ and $\bar{P}_i$, $\bar{x}_i$, are the average values for clones (for region i) and region i respectively. This price index is called the Regional Fisher Housing Price Index. This measurement removes the bias arising from different housing baskets through the aforementioned procedure to build the samples. Finally, the geometric mean represents the price index between region i and benchmark region mr. Nonetheless, the regional housing price index must be constructed among all regions, and not only between each one of the regions $i = 1, 2, ..., 12$ and mr. The Fisher Price Index allows the direct comparison among regions using the superlative propriety (Diewert, 1978). The Spatial Fisher Housing Price Index between region i and region $j \in i = 1, 2, ..., 12 \forall i \neq j$ is given by the quotient between the price index of the region i and j in respect to mr. This property can be established as

$$IF_{i/j} = \frac{IF_{i/mr}}{IF_{j/mr}} \tag{4.6}$$

With this information, this paper reports a matrix of the regional housing price index for the 12 regions, plus the benchmark region.

$$MIF = \begin{bmatrix} 1 & IF_{1/2} & ... & IF_{1/mr} \\ ... & ... & ... & ... \\ IF_{mr/1} & IF_{mr/2} & ... & 1 \end{bmatrix}$$

The cell identified in the first row and second column represents the index between regions 1 and 2. If this number is greater than 1, it means that an individual that moves from 1 to 2 will spend, on average, an additional percentage of rent equal to $(IF_{1/2} - 1)$

4.4 The data

The data are extracted from the Chilean Household Survey 2006 (CASEN 2006). This sample consists of 268,873 individuals covering 13 regions: 12 regions, plus the Metropolitan Region. The latter will be used as a benchmark because it offers more observations and more probability of finding a clone for regional houses. Figure 1 describes the geographic distribution of the 12 regions (IXII) and Metropolitan Region (RM). Hedonic regressions require the price and characteristics of the houses. CASEN 2006 provides this information for 7,184 householders who rent houses. Outliers and missing data[4] were eliminated, leaving 7,094 observations. CASEN 2006 includes a weight variable for each observation; therefore the sample can be weighted to the population. Such as Table 4.1 shows, this implies that the 7,094 householders are expanded to 649.328[5]

4.5 Results

4.5.1 Testing results for matching estimator methods

Regional housing differences can be measured in two dimensions: covariates and propensity score. In the first case, two houses are different if their characteristics are different. Some comparability problems can appear because there are several covariates. Therefore, the differences can exist in some variables and not in the others. This problem can be overcome by using the single-measure propensity score. This number

[4]Any head of householder without information about the price of the house, household characteristics or characteristics of the householder was considered a missing value. This is the best selection given the data. Possibly, If the survey had information about the price paid by house owners, the index could have been more representative. However, it was not the case, and this index is a good representation given the available information.

[5]The CASEN survey provides information about housing characteristics, but its quality is limited. There is no information about important physical characteristics, such as the size in square meters, age, and whether there is a garage. As well, there is no data for spatial location, for example, distance to the central business district or place of work or proximity to services centers. These limitations restrict the available models used to estimate the regional housing price index.

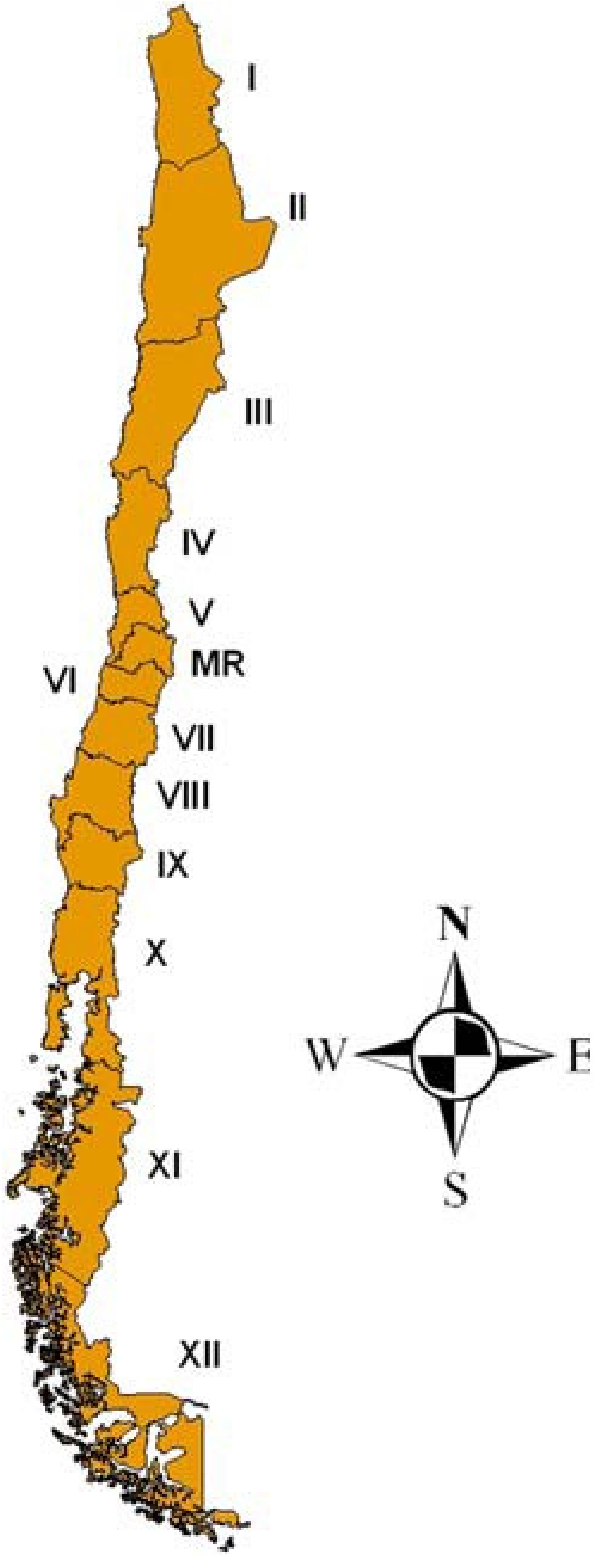

Figure 4.1: Map of Chile.

Table 4.1: Covariates

Variables [a]	**Description**	**As used for original covariate**	**Mean estimating the propensity score**	**SD**	**Min**	**Max**
Bedroom	Number	Integers	2.4	1.04	0	12
Alternative use room	Number	Integers	0.86	0.47	0	3
Bathroom	Number	Integers	1.08	0.49	0	5
Kitchen	Number	Integers	0.81	0.41	0	3
Housing quality	Index quality	Index between 0 and 1	88.55	9.3	28.53	99.97
Water heater	Dummy	1 = yes, 0 = no	0.61	0.49	0	1
Telephone	Dummy	1 = yes, 0 = no	0.28	0.45	0	1
Cable TV	Dummy	1 = yes, 0 = no	0.28	0.45	0	1
Crowding	Number	Persons per bedroom	1.59	0.82	0.09	8.00
Education	Household head' eduaction	Integers, years of education	11.08	3.95	0	20
Age	Years	Years	41.29	12.82	16	98
Marital status	Dummy	1 = married, 0 = notmarried	0.45	0.50	0.00	1.00
Sex	Dummy	1 = man, 0 = woman	0.75	0.43	0.00	1.00

[a] Descriptive statistics for 7,094 observations

represents the probability of a householders living in a specific region for treated and non treated houses, therefore a similar propensity score implies similar houses under the assumptions detailed by Rosembaum and Rubin (1985). This paper uses both measurements to prove the adjustment of matching methods.

Table 4.2: Sample means of covariate for CASEN 2006 by region

Region	**Weighted sample**	**No. of obs. sample**	**Housing rent**	**Bedroom**	**Alt. use room**	**Bathroom**	**Kitchen**	**Quality housing**	**Crowding**	**Education**	**Age**
I	22,858	214	11.10 (0.52)	2.29 (1.25)	0.75 (0.55)	0.99 (0.44)	0.74 (0.46)	79.44 (14.8)	1.74 (0.91)	11.06 (3.75)	41.28 (12.8)
II	22,838	293	11.26 (0.59)	2.11 (1.00)	0.79 (0.49)	1.06 (0.41)	0.77 (0.47)	87.71 (9.52)	1.91 (1.11)	11.53 (3.64)	39.18 (12.3)
III	9,880	195	10.95 (0.52)	2.39 (0.89)	0.87 (0.45)	1.01 (0.33)	0.84 (0.37)	90.26 (7.38)	1.61 (0.84)	11.26 (3.62)	39.52 (11.1)
IV	15,642	177	10.93 (0.58)	2.47 (1.16)	0.78 (0.44)	1.07 (0.51)	0.76 (0.42)	87.57 (9.81)	1.53 (0.69)	10.92 (4.12)	41.40 (13.7)
V	69,050	854	11.08 (0.51)	2.46 (0.91)	0.94 (0.39)	1.11 (0.42)	0.91 (0.30)	92.39 (5.46)	1.56 (0.75)	11.26 (3.72)	42.81 (13.4)
VI	27,333	522	10.82 (0.49)	2.43 (0.97)	0.93 (0.41)	0.98 (0.38)	0.88 (0.33)	82.95 (12.4)	1.59 (0.78)	10.11 (3.87)	41.46 (12.6)
VII	24,592	357	10.75 (0.52)	2.41 (1.00)	0.87 (0.48)	1.03 (0.42)	0.79 (0.40)	84.59 (12.2)	1.60 (0.88)	10.11 (4.21)	42.48 (13.0)
VIII	60,946	904	10.87 (0.55)	2.34 (0.87)	0.86 (0.47)	1.02 (0.42)	0.78 (0.43)	85.95 (9.32)	1.61 (0.80)	11.03 (3.83)	40.78 (12.9)
IX	24,353	415	10.78 (0.49)	2.37 (0.86)	0.84 (0.50)	1.03 (0.35)	0.72 (0.45)	87.47 (7.66)	1.52 (0.77)	10.73 (4.13)	40.85 (13.2)
X	36,297	693	10.98 (0.50)	2.48 (1.12)	0.71 (0.50)	1.00 (0.37)	0.64 (0.48)	89.90 (8.17)	1.55 (0.78)	10.81 (3.95)	39.07 (11.5)
XI	4,706	174	11.30 (0.52)	2.61 (0.90)	0.70 (0.52)	1.01 (0.31)	0.63 (0.48)	93.68 (6.05)	1.41 (0.67)	11.42 (3.86)	38.99 (10.7)
XII	5,504	127	11.30 (0.49)	2.43 (0.96)	0.77 (0.47)	1.03 (0.56)	0.66 (0.47)	79.53 (15.5)	1.57 (0.88)	10.60 (4.41)	41.50 (11.5)
MR	325,329	2,169	11.42 (0.58)	2.41 (1.14)	0.91 (0.45)	1.20 (0.60)	0.89 (0.35)	90.96 (5.99)	1.59 (0.82)	11.52 (3.99)	42.08 (13.0)
Total	649,328	7,094									

Table 4.2 provides the baseline to analyze the regional heterogeneity of housing, using mean and standard deviations. The first 12 rows represent each region and row 13, the Metropolitan area (benchmark). The first column has the designated name for each region, which is used in this paper to refer to the respective region. The second and third columns are the sample and weighted sample by region, respectively. The Housing Rent column is the regional average of the natural logarithm of the housing price. Finally, the rest of the columns are variables used to estimate the three matching estimator methods. Table 4.2 shows different average values for each of the 13 regions. For example, the 11th Region has a mean of 2.61 bedrooms per house, which is different from the 2.11 of the second Region. These differences are evident for almost all variables, specifically in Housing Quality, showing the presence of the geographic heterogeneity across the regions. These differences are also found in the standard deviations. For example, the variable "bedroom" has a standard deviation of 0.86 in the ninth Region and 1.25 in the first Region.

Table 4.3: Standardized difference by region baseline

Region	**Bedroom difference**	**Alternative use room**	**Bathroom**	**Kitchen**	**Quality housing**	**Crowding**	**Education**	**Age**	**Pscore**
I	**-9.84**	-31.91	-39.53	-34.64	-101.58	16.25	-11.85	**-6.21**	0.24
II	-28.05	-25.91	-26.75	-26.86	-40.84	31.97	**0.32**	-22.91	0.06
III	**-1.74**	**-9.37**	-38.06	-14.02	-10.52	**1.64**	**-6.92**	-21.08	0.01
IX	**5.82**	-30.05	-22.00	-31.69	-41.73	**-8.39**	-14.76	**-5.15**	0.04
V	**4.63**	**5.32**	-17.46	**6.01**	24.84	**-4.42**	**-6.65**	**5.48**	0.05
VI	**2.57**	**3.86**	-42.65	**-3.23**	-81.98	**-0.34**	-35.92	**-4.86**	0.18
VII	**0.39**	**-9.80**	-31.69	-25.38	-66.15	**0.34**	-34.28	**3.04**	0.11
VIII	**-6.57**	-11.11	-33.50	-28.27	-63.96	**2.52**	-12.55	-10.08	0.16
IX	**-3.61**	-15.27	-32.73	-42.23	-50.78	**-9.68**	-19.43	**-9.41**	0.09
X	**6.07**	-42.00	-39.02	-58.57	-14.85	**-5.93**	-17.88	-24.51	0.12
XI	19.59	-44.83	-39.60	-61.40	45.06	-23.76	**-2.56**	-25.86	0.15
XII	**2.42**	-30.72	-28.28	-53.87	-96.97	**-2.43**	-21.89	**-4.78**	0.22

Bold numbers are less than 10% in absolute value except for Pscore difference

Table 4.3 shows the standardized difference for the regional mean sample. This statistical measurement is comparable without taking the metrics of each variable into account and represents the difference between means as a percentage of the variance. The covariates for the matching estimator are between the second and eighth columns. The last column is the average propensity score between each region and the Metropolitan Region. The greatest differences are between extreme regions (i.e. the first and 12th Regions I) and the Metropolitan Region (central). According to Rosembaum and Rubin (1985)), standardized differences under 10% are acceptable to make comparisons among the regions. In this case, a 37.5% difference would be acceptable to make comparison among regions.

The first test is estimated for NNM matching method and the regional means are displayed in Table 4.4. It shows a significant reduction of the differences with respect to the baseline. Indeed, 64% of the averages are

Table 4.4: Standardized difference by region nearest neighbor

Region	Bedroom	Alternative use room	Bathroom	Kitchen	Quality housing	Crowding	Education	Age	Pscore difference
I	-15.54	25.74	-13.48	41.48	-49.56	36.25	-46.14	15.09	0.05
II	**-3.24**	**3.37**	**3.54**	**-0.75**	**1.43**	**-3.21**	**5.47**	**-1.22**	0.00
III	**-2.96**	**4.62**	**-7.55**	**4.07**	**-1.45**	**-4.31**	**-2.93**	**8.44**	0.00
IV	**-4.93**	**4.66**	**-8.59**	**1.28**	**0.59**	10.83	**7.69**	-12.16	0.00
V	-27.51	-11.78	**0.00**	**-7.28**	14.40	17.77	**7.22**	-12.96	0.01
VI	-25.05	-11.15	**8.58**	**-3.88**	-40.69	25.20	**0.64**	18.95	0.07
VII	**-7.52**	**6.54**	**1.64**	11.80	-16.49	11.84	**2.33**	**5.59**	0.02
VIII	**-1.73**	**1.72**	**1.55**	**9.35**	-45.68	11.65	-25.01	20.71	0.07
IX	**2.36**	**-0.44**	**3.24**	**5.27**	**-8.37**	**5.12**	**-4.40**	14.09	0.00
X	**7.32**	**0.55**	**2.14**	**-4.48**	**-0.33**	**-5.07**	**-0.48**	**-0.39**	0.01
XI	**-2.55**	21.46	**2.21**	22.77	34.39	**-1.40**	13.83	**-1.66**	0.03
XII	**-4.37**	-11.63	**5.59**	**-5.02**	-18.10	**1.78**	-17.38	**8.79**	0.05

Bold numbers are less than 10% in absolute value except for Pscore difference

Table 4.5: Standardized difference by region Mahalanobis - propensity score covariate

Region	Bedroom	Alternative use room	Bathroom	Kitchen	Quality housing	Crowding	Education	Age	Pscore difference
I	**0.73**	**0.00**	10.48	**-2.28**	-43.82	**-1.48**	**3.88**	12.30	0.09
II	**0.91**	**-0.92**	**2.11**	**0.97**	**-4.57**	**1.41**	**0.12**	**0.04**	0.01
III	**0.74**	**0.00**	**0.00**	**0.00**	**6.36**	**6.52**	**-2.49**	**3.04**	0.00
IV	**4.37**	**1.53**	**2.80**	**0.00**	**1.48**	**-0.87**	**-2.96**	**2.92**	0.00
V	**-0.32**	**0.00**	**2.68**	**0.00**	**2.69**	**7.00**	**-5.55**	**0.03**	0.01
VI	**4.61**	**1.85**	**1.10**	**-1.13**	-28.52	**-1.18**	**-6.82**	10.34	0.04
VII	**3.51**	**-0.97**	**1.08**	**-2.54**	**-6.05**	**-3.31**	**1.85**	**2.92**	0.01
VIII	**3.46**	**0.67**	**1.48**	**-0.91**	-24.80	**1.04**	**-5.58**	10.69	0.05
IX	**7.82**	**0.97**	**4.84**	**-2.14**	-14.16	**-5.20**	**-0.42**	**3.49**	0.02
X	**4.77**	**-1.34**	**0.00**	**-0.79**	13.29	**-2.56**	**-0.92**	**-1.08**	0.01
XI	**1.34**	**0.00**	**-9.83**	**0.00**	**-5.57**	**-0.83**	**-7.36**	**1.34**	0.01
XII	**4.29**	**0.00**	**0.00**	**-3.19**	**-6.29**	**-7.69**	12.81	**-2.94**	0.02

Bold numbers are less than 10% in absolute value except for Pscore difference under 10%. Nevertheless, there are some variables with persistent differences, for example housing quality variables. The bias measured as a difference in propensity score exhibits a significant narrowing in relation to the base line.

According to Table 4.5, the MMPS fits better than the NNM. The MMPSs heterogeneity reduction is greater than that of the NNM, reaching 89.58%. Using the one-dimensional variable (propensity score), there is a marginal improvement in the NNM model. Nevertheless, the MMPS could not reduce the bias for the first Region, where the difference propensity score is 0.09. In any case, this model considerably reduces the geographic bias with respect to the baseline compared to the NNM. Table 4.6 shows the output for the MMWPS method. This method reduces heterogeneity more than the other two matching methods. The reduction is 90.63% compared to the baseline, removing completely the bias in alternative use room, bathroom and kitchen. On the other hand, the MMWPS reduces the bias in housing quality, although not to less than 10% in some regions. Nevertheless, the average improvement is significant. This method provides the best fit, reaching 90.63%, better than the 64% and 89.58% for the other two methods using all of variables, and reducing the bias in the propensity score. The summary for these models is presented in the Table 4.7. The best method is the MMWPS. This method reduces the difference in propensity score and covariates. The comparison between the first and fourth column indicates the improvement achieved by using different matching estimator methods.

Table 4.6: Standardized difference by region Mahalanobis-propensity score covariate (caliper 0.2 x variance Pscore)

Region	Bedroom	Alternative use	Bathroom	Kitchen	Quality housing	Crowding	Education	Age	Pscore difference
I	**-2.80**	**0.00**	**3.52**	**-3.48**	-42.77	**0.72**	12.29	11.50	0.06
II	**0.91**	**-0.92**	**2.11**	**0.97**	**-4.57**	**1.41**	**0.12**	**0.04**	0.01
III	**0.74**	**0.00**	**0.00**	**0.00**	**6.36**	**6.52**	**-2.49**	**3.04**	0.00
IV	**4.37**	**1.53**	**2.80**	**0.00**	**1.48**	**-0.87**	**-2.96**	**2.92**	0.00
V	**-0.31**	**0.00**	**2.66**	**0.00**	**1.44**	**7.78**	**-5.58**	**0.14**	0.01
VI	**7.24**	**2.25**	**0.00**	**0.00**	-15.33	**-8.59**	**2.70**	**5.52**	0.02
VII	**2.66**	**0.00**	**1.12**	**-1.28**	**-5.31**	**-3.06**	**3.18**	**4.73**	0.01
VIII	11.28	**0.87**	**4.58**	**-1.15**	-17.29	**-1.14**	**0.82**	**5.90**	0.03
IX	**7.82**	**0.97**	**4.84**	**-2.14**	-14.16	**-5.20**	**-0.42**	**3.49**	0.02
X	**3.77**	**-1.45**	**-0.84**	**0.00**	10.95	**-1.53**	**-1.92**	**-1.10**	0.01
XI	**1.34**	**0.00**	**-9.83**	**0.00**	**-5.57**	**-0.83**	**-7.36**	**1.34**	0.01
XII	**4.29**	**0.00**	**0.00**	**-3.19**	**-6.29**	**-7.69**	12.81	**-2.94**	0.02

Bold numbers are less than 10% in absolute value except for Pscore difference

4.6 Results for hedonic regression and price index matrix

Each region has two groups of hedonic prices. The first group was computed using the housing weight data available to each region and the second represents the same methodology applied to each regional control group belonging to the Metropolitan Region. The results are provided in Tables 4.8 and 4.9. The variables used represent characteristics of the head of the household and household attributes[6]. The upper level of the table contains the estimation with regional data. Most of the variables show the expected sign. Education shows a positive coefficient and is significant for most of the regions, supporting the hypothesis that people obtain better quality houses if they achieve higher education levels. Marital status and sex has an irregular behavior throughout the regions, showing that there is little evidence supporting significant differences in housing. On the other hand, the "housing quality" variable has the expected positive sign for all the regions, except the First Region. Finally, the measure of fit of the model lies between 0.49 and 0.85.

Table 4.7: Summary of matching methods

Region/method	**PS base line**	**PS NNM**	**PS MMPS**	**PS MMWPS**
I	0.24	0.05	0.09	0.06
II	0.06	0.00	0.01	0.01
III	0.01	0.00	0.00	0.00
IV	0.09	0.00	0.02	0.02
V	0.04	0.00	0.00	0.00
VI	0.05	0.01	0.01	0.01
VII	0.18	0.07	0.04	0.02
VIII	0.11	0.02	0.01	0.01
IX	0.16	0.07	0.05	0.03
X	0.12	0.01	0.01	0.01
XI	0.15	0.03	0.01	0.01
XII	0.22	0.05	0.02	0.02
Variable differences (%)	37.50	64.00	89.58	90.63

The lower part of the table reflects the same hedonic regression methodology, but applied to the control group. One of the advantages of this methodology is that it allows comparison in hedonic price estimation for "comparable houses". In this sense, the hedonic prices are comparable because they come from comparable samples, reducing quality bias among regions. In general, the adjustment for the second group is similar to the first.

The Fisher index considers both bundles, for the region and Metropolitan Region, reducing the geographical bias and allowing comparability across regions. Table 4.10 contains the Regional Housing Price Index for Chilean regions. Each cell along the rows shows the price index between the region in column j and the one along row i , that has been set at the numerarie equal to one. For example, the housing cost between the first Region and the second Region is 1.22. This indicates that the housing

[6]This estimation could present problems of endogeneity between the price of the house and the level of income. Given this problem, control of the household characteristics could improve the consistency of the estimations. This same problem could be argued for the endogeneity generated by spatial dependence, but this problem cannot be treated with the available information (Bowden, 1992)

Table 4.8: Hedonic regression for treatment and control housing

Variable/region	**I**	**II**	**III**	**IV**	**V**	**VI**	**VII**	**VIII**	**IX**	**X**	**XI**	**XII**
Treatment												
Constant	11.64 (92.38)	7.75 (72.42)	9.42 (73.34)	7.88 (87.72)	9.06 (118.7)	7.81 (90.32)	9.05 (188.6)	8.57 (111.8)	8.40 (153.2)	9.38 (103.4)	10.36 (50.07)	10.03 (127.3)
Education	0.03 (12.19)	0.03 (19.60)	0.04 (26.30)	0.03 (19.13)	0.02 (16.12)	0.02 (13.46)	0.03 (28.03)	0.02 (15.36)	0.00 (-2.69)	0.03 (22.75)	0.02 (8.966)	0.05 (28.38)
Marital status	0.29 (26.07)	-0.04 (-6.33)	0.18 (18.86)	-0.16 (-18.7)	-0.01 (-2.62)	0.08 (9.905)	-0.03 (-5.78)	0.00 (0.626)	-0.11 (-16.9)	-0.04 (-6.10)	-0.02 (-1.57)	0.05 (6.525)
Sex	0.05 (4.831)	-0.02 (-2.08)	0.02 (1.294)	0.12 (13.56)	0.00 (0.236)	-0.02 (-2.47)	-0.06 (-10.4)	-0.04 (-5.29)	0.05 (6.707)	0.01 (1.527)	0.03 (1.387)	0.17 (19.27)
Age	0.00 (3.254)	0.00 (-6.54)	0.01 (15.28)	0.00 (9.289)	0.00 (7.573)	0.01 (26.20)	0.00 (8.380)	0.00 (-2.39)	0.00 (-3.94)	0.00 (-8.71)	0.01 (15.31)	0.00 (1.247)
Bedroom	0.06 (14.62)	0.06 (16.14)	0.06 (10.48)	-0.06 (-13.7)	0.10 (38.09)	0.10 (24.74)	0.15 (51.13)	0.05 (13.10)	0.09 (25.12)	0.09 (38.02)	0.00 (0.063)	0.24 (48.68)
Alt. use room	-0.01 (-0.68)	0.12 (9.648)	-0.06 (-3.39)	-0.10 (-6.76)	-0.16 (-18.0)	-0.02 (-1.49)	0.04 (6.078)	0.16 (21.59)	0.03 (4.393)	0.13 (15.72)	0.22 (11.26)	-0.18 (-16.6)
Bathroom	-0.13 (-10.1)	0.17 (25.44)	0.46 (34.93)	0.37 (37.50)	0.43 (93.62)	0.44 (47.92)	0.32 (48.92)	0.32 (47.09)	0.35 (45.80)	0.26 (34.75)	0.64 (27.69)	0.10 (13.80)
Kitchen	0.45 (30.00)	-0.12 (-9.10)	0.16 (7.543)	0.17 (11.65)	0.12 (12.28)	0.23 (16.72)	0.12 (13.90)	0.01 (0.936)	0.08 (8.460)	-0.08 (-8.44)	-0.28 (-9.86)	-0.07 (-6.31)
Quality	-0.02 (-8.93)	0.03 (24.72)	0.00 (-0.62)	0.02 (21.81)	0.01 (10.67)	0.02 (15.70)	0.00 (7.767)	0.02 (16.67)	0.02 (29.29)	0.01 (8.549)	-0.01 (-2.87)	0.00 (-2.07)
Water heater	0.44 (43.92)	0.27 (39.52)	0.25 (23.50)	0.47 (51.72)	0.27 (36.89)	0.20 (21.26)	0.33 (49.81)	0.25 (33.43)	0.28 (30.66)	0.37 (51.73)	0.36 (16.33)	0.13 (9.925)
Phone	0.13 (11.07)	0.13 (17.58)	0.11 (10.44)	0.26 (24.66)	0.04 (7.117)	-0.02 (-1.69)	0.05 (6.665)	0.16 (21.74)	0.18 (24.57)	0.21 (25.60)	-0.12 (-7.27)	0.24 (25.97)
Cable TV	-0.04 (-4.26)	0.01 (0.954)	0.11 (10.16)	-0.07 (-6.30)	0.20 (40.53)	0.10 (12.83)	0.18 (24.92)	0.19 (25.63)	0.18 (24.75)	0.03 (4.625)	0.30 (19.92)	0.15 (16.57)
Fit of measure	0.53	0.49	0.64	0.53	0.58	0.63	0.71	0.64	0.65	0.63	0.77	0.85

Table 4.9: Hedonic regression for treatment and control housing

Variable/region	**I**	**II**	**III**	**IV**	**V**	**VI**	**VII**	**VIII**	**IX**	**X**	**XI**	**XII**
Control												
Constant	10.24 (60.64)	8.40 (75.86)	8.98 (72.58)	8.18 (112.6)	8.28 (87.88)	9.00 (103.9)	9.27 (189.6)	9.76 (123.6)	9.06 (129.7)	10.58 (113.4)	7.35 (26.69)	9.62 (72.64)
Education	0.02 (8.325)	0.05 (33.01)	0.02 (12.58)	0.02 (18.77)	0.01 (3.880)	0.02 (12.48)	0.02 (19.28)	0.04 (34.16)	0.03 (17.65)	0.08 (58.62)	0.00 (0.742)	0.08 (27.29)
Marital status	0.56 (55.36)	-0.09 (-14.7)	-0.22 (-19.5)	-0.01 (-1.34)	-0.05 (-8.92)	-0.05 (-5.89)	-0.02 (-2.91)	-0.09 (-13.2)	0.12 (13.93)	-0.13 (-19.2)	-0.08 (-6.18)	0.33 (20.22)
Sex	-0.35 (-30.7)	-0.08 (-11.9)	-0.06 (-5.39)	-0.02 (-2.51)	0.06 (9.241)	-0.17 (-17.6)	0.00 (-0.22)	-0.05 (-7.68)	0.09 (8.099)	0.06 (8.269)	0.09 (6.774)	-0.09 (-7.74)
Age	0.00 (8.574)	0.01 (32.57)	0.00 (1.257)	0.01 (45.29)	0.00 (19.80)	0.00 (13.96)	0.00 (9.170)	0.01 (33.29)	0.00 (5.025)	0.01 (22.18)	0.00 (-5.57)	0.03 (42.79)
Bedroom	-0.06 (-14.0)	-0.02 (-5.28)	0.13 (20.63)	-0.02 (-5.26)	-0.02 (-4.57)	0.03 (7.626)	0.08 (25.99)	0.06 (15.31)	0.07 (14.07)	-0.04 (-13.7)	0.01 (1.547)	-0.04 (-4.32)
Alt. use room	0.15 (22.63)	0.37 (30.51)	0.15 (8.341)	-0.04 (-3.60)	0.46 (45.44)	0.14 (13.05)	0.21 (29.69)	0.03 (3.754)	0.02 (1.545)	-0.02 (-1.86)	0.20 (12.35)	0.39 (25.94)
Bathroom	0.49 (32.36)	0.40 (61.46)	0.07 (4.829)	0.45 (53.17)	0.43 (75.52)	0.56 (61.60)	0.28 (43.68)	0.30 (41.10)	0.21 (20.79)	0.34 (43.24)	0.61 (30.40)	0.35 (31.22)
Kitchen	-0.20 (-13.6)	0.03 (2.115)	-0.12 (-5.48)	0.25 (20.56)	-0.15 (-13.3)	0.11 (8.049)	-0.07 (-7.81)	-0.07 (-6.18)	0.08 (5.844)	0.11 (11.00)	-0.19 (-7.93)	-0.16 (-11.9)
Quality	0.00 (1.644)	0.01 (7.652)	0.02 (10.64)	0.02 (19.81)	0.02 (18.13)	0.01 (10.47)	0.01 (15.47)	0.00 (1.819)	0.01 (13.78)	-0.01 (-8.27)	0.04 (11.30)	-0.01 (-7.04)
Water heater	0.11 (8.996)	0.30 (42.00)	0.28 (21.36)	0.17 (21.70)	0.19 (24.34)	0.21 (25.16)	0.12 (16.41)	0.19 (24.37)	0.38 (34.55)	0.13 (16.21)	-0.07 (-4.04)	0.10 (6.691)
Telephone	0.11 (10.03)	0.10 (14.38)	0.20 (18.19)	0.17 (24.09)	0.24 (36.32)	0.16 (18.26)	0.16 (24.75)	0.14 (20.77)	-0.11 (-11.7)	0.22 (30.72)	0.29 (20.40)	0.48 (30.17)
Cable TV	0.27 (18.21)	0.16 (22.65)	0.02 (1.705)	0.11 (13.88)	-0.06 (-9.88)	-0.07 (-7.75)	0.39 (50.30)	0.19 (26.61)	0.29 (29.29)	0.04 (4.441)	0.20 (14.55)	0.00 (0.131)
Measure of -t	0.61	0.66	0.48	0.61	0.56	0.63	0.69	0.56	0.53	0.48	0.77	0.87

Prices are higher in the second Region by 22%. The matrix shows different results compared to the fourth column in Table 4.2, "Housing Rent". Considering only the average house price in the region, the metropolitan Region has the most expensive values in the country. According to Table 4.10, differential regional prices have a different behavior when heterogeneity is considered in the estimation process. Particularly, the most expensive region in housing is the second, where the maximum difference reaches 73% and is between the second and seventh Regions.

Table 4.10: Matrix of regional housing price index

Region	**I**	**II**	**III**	**IV**	**V**	**VI**	**VII**	**VIII**	**IX**	**X**	**XI**	**XII**	**MR**
I	1.00	1.22	0.81	0.76	0.80	0.74	0.71	0.84	0.75	0.84	0.94	0.87	1.03
II	0.82	1.00	0.66	0.63	0.65	0.61	0.58	0.69	0.62	0.69	0.77	0.72	0.85
III	1.24	1.51	1.00	0.95	0.99	0.92	0.87	1.04	0.93	1.04	1.16	1.08	1.28
IV	1.31	1.60	1.06	1.00	1.04	0.97	0.92	1.09	0.98	1.10	1.23	1.14	1.35
V	1.26	1.53	1.01	0.96	1.00	0.93	0.89	1.05	0.94	1.05	1.18	1.10	1.30
VI	1.35	1.64	1.09	1.03	1.07	1.00	0.95	1.13	1.01	1.13	1.27	1.18	1.39
VII	1.42	1.73	1.14	1.08	1.13	1.05	1.00	1.18	1.06	1.19	1.33	1.24	1.47
VIII	1.20	1.46	0.97	0.91	0.95	0.89	0.84	1.00	0.90	1.00	1.12	1.05	1.24
IV	1.33	1.62	1.07	1.02	1.06	0.99	0.94	1.11	1.00	1.11	1.25	1.16	1.38
X	1.19	1.45	0.96	0.91	0.95	0.89	0.84	1.00	0.90	1.00	1.12	1.04	1.23
XII	1.07	1.30	0.86	0.81	0.85	0.79	0.75	0.89	0.80	0.89	1.00	0.93	1.10
XII	1.14	1.39	0.92	0.87	0.91	0.85	0.81	0.96	0.86	0.96	1.07	1.00	1.18
MR	0.97	1.18	0.78	0.74	0.77	0.72	0.68	0.81	0.73	0.81	0.91	0.85	1.00

4.7 Policy implications

Estimation of the Regional Housing Price Index has at least three policy implications for using the index as a proxy for the regional cost of living: correction of regional economic variables, regional wage bargaining and regional poverty rates. The Regional Housing Price index can be considered as the first regional measurement of the regional cost of living in Chile. Considering that housing data is the only regional information available on prices and consumption, this information is useful as a proxy to estimate the regional cost of living.With this information, it will now be possible to re-estimate the regional economic variable (GDP) by incorporating more regionally specific price information. Secondly, this measure will facilitate regional wage bargaining policies using the evidence for the spatial heterogeneity of the cost of living (Aroca and Atienza, 2008). Problems could potentially arise in the national bargaining process, since the income required to maintain consumers at the same level of utility is not the same throughout the country.

Thirdly, using the proposed index, the government could re-calculate the poverty rate for each region. Social policy in Chile is currently decided in consideration of a homogeneous poverty line for each region. However, the homogeneous pricing method hides important effects that price differences may have at the regional level. Thus, the correction in determining regional poverty rates could help to focus social spending more appropriately throughout the country.

4.8 Conclusion

This paper computes a regional housing price index using quasi-experimental control group methods, hedonic regression and a spatial price index. The adoption of quasi-experimental control group methods reduces geographic heterogeneity in the regional comparison of housing. Subsequently, a matrix of the regional housing price index is reported using hedonic price and spatial Fisher price indice. This methodology improves the results obtained from only averaging regional housing prices. The application of the methodology to Chilean data shows dramatic changes in the results. Different results are maintained for any pair of regions in the matrix, indicating the varying results of simple average price on the one hand, and a regional housing price index that considers geographical heterogeneity on the other. In particular, this paper tested three control group methods: (1) nearest matching on the propensity score; (2) Mahalanobis matching including the propensity score; and (3)Mahalanobis matching within score calipers. The evaluation of these methods applied to Chilean data shows that the third method provides the greatest reduction of average regional bias measured through standardized differences. Finally, this paper shows that the Regional Housing Price Index must take geographical heterogeneity into account. Otherwise, as in the case of the average price methodology, it can generate an inaccurate picture of the regional cost of housing and therefore provide incorrect signals to national and regional government agencies whose policy initiatives center on reduction of regional disparities.

Chapter 5

Conclusion

These three essays contribute to an evaluation of the sources of the spatial concentration of wages and what are their potential consequences on this spatial distribution. Assuming that the excessive spatial concentration of wages has several negative consequences on equity and efficiency, then it is very important to identify with a causal mechanism (NEG) those forces. The results show that NEG is not sufficient for understanding the spatial wage distribution in a developing country with high dependency of natural resources. Moreover, alternative frameworks with the incorporation of natural resources help to evaluate the case. Simultaneously, the spatial concentration process must be encompassed with role of congestion cost, such as housing prices.

This thesis opens a new discussion in the literature with a special call for economists to incorporate spatial economic modeling, but where the spatial location of initial endowment play a fundamental role in the final equilibrium.

Chapter 6

References

Alonso W. 1964. *Location and land use.* Cambridge: MA: Harvard University Press.

Armstrong H, Taylor J. 2000. *Regional Economics and Policy.* Oxford: Blackwell Publishers.

Aroca P. 2009. Las desigualdades regionales en chile. *Foreign Affairs* **9**: 53–62.

Aroca P, Atienza M. 2008. La conmutacin regional en chile y su impacto en la regin de antofagasta. *EURE* **102**: 97–120.

Aroca P, Bosch M. 2000. Crecimiento, convergencia y espacio en las regiones chilenas: 1960-1998. *Estudios de Economia* **27**: 199–224.

Aroca P, Hewings G. 2002. Migration and regional labor market adjustment: Chile 1977-1982 and 1987-1992. *Annals of Regional Science* **36**: 197–6.

Atkin M, Longford N. 1986. Statistical modelling in school effectiveness studies. *Journal of the Royal Statistical Society* **149**: 1–43.

Baldwin R, Forslid R, Martin P, Ottaviano G, Robert-Nicoud F. 2003. *Economic Geography and Public Policy.* New Jersey: Princeton University Press.

Bowden R. 1992. Competitive selection and market data: the mixed-index problem. *Review of Economic Studies* **59**: 625–633.

Brakman S, Garretsen H, Gorter J, van der Horst A, Schramm M. 2005. New economic geography, empirics, and regional policy.
URL `"http://ideas.repec.org/p/cpb/spcial/56.html"`

Brakman S, Garretsen H, Schramm M. 2004. The spatial distribution of wages: estimating the Helpman-Hanson model for Germany. *Journal of regional Science* **44**: 437–466.

Brakman S, Garretsen H, Schramm M. 2006. Putting new economic geography to the test: free-ness of trade and agglomerations in the eu regions. *Regional Science and Urban Economics* **35**: 613–635.

Brakman S, Garretsen H, van Marrewijk C. 2009. *The New Introduction to Geographical Economics.* Cambridge: Cambridge University Press.

Breinlich H. 2006. The spatial income structure in the european union-what role for economic geography? *Journal of Economic Geography* **6**: 593–617.

Brulhart M, Koenig P. 2006. New economic geography meets comecon: regional wages and industry location in central europe. *Economics of Transition* **14**: 245–267.

Cochran W, Rubin D. 1973. Controlling bias in observational studies: A review. *Sankhya* **35**: 417–446.

Combes P, Duranton G, Gobillon L. 2008a. Spatial wage disparities: Sorting matters! *Journal of Urban Economics* **63**: 732–742.

Combes P, Lafourcade M, Thisse J, Toutain J. 2008b. The rise and fall of spatial inequalities in france: a long-run perspective.
URL `http://www.cepr.org/pubs/dps/DP7017.asp`

Contreras D. 2003. Poverty and inequality in a rapid growth economy: Chile 1990-96. *Journal of Development Studies* **39**: 181–200.

Costello G, Watkins C. 2002. Towards a system of local house price indices. *Housing Studies* **17**: 857–873.

Diewert W. 1978. Superlative index numbers and consistency in aggregation. *Econometrica* **46**: 883–900.

Dixit A, Stiglitz J. 1977. Monopolistic competition and optimum product diversity. *American Economic Review* : 297–308.

Dumond J, Hirsch B, Macpherson D. 1999. Wage differentials across labor markets and workers: does cost of living matter? *Economic Inquiry* **37**: 577–598.

Duncan R, Fuentes J. 2006. Convergencia regional en chile: nuevos test, viejos resultados. *Cuadernos de Economia* **43**: 81–112.

Echeverria R, Gopinath M. 2007. *Geographic concentration of Economic Actvity: The case of Chile.*
URL `http://aede.osu.edu/Programs/Anderson/trade/65EcheverriaGopinath.pdf`

Fally T, Paillacar R, Terra C. 2010. Economic geography and wages in brazil: Evidence from micro-data. *Journal of Development Economics* **91**: 155–168.

Ferreira F, Litchfield J. 1999. Calm after the storms: income distribution and welfare in chile 1987-1994. *World Bank Economic Review* **13**: 509–538.

Fingleton B. 2005. Beyond neoclassical orthodoxy: a view based on the new economic geography and uk regional wage data. *Papers in Regional Science* **84**: 351–375.

Fingleton B. 2006. The new economic geography versus urban economics: an evaluation using local wage rates in Great Britain. *Oxford Economic Papers* **58**: 501–530.

Fingleton B. 2008. Competing models of global dynamics: evidence from panel models with spatially correlated error components. *Economic Modelling* **25**: 542–558.

Fingleton B. 2009. Simulating wages and house prices using the neg. *Discussion papers in Economics University of Strachclyde.* **09-13**.

Forret D. 1991. An analysis of house price differentials between english regions. *Regional Studies* **25**: 231–238.

Fujita M, Krugman P, Venables A. 1999. *The Spatial Economy: Cities, Regions, and International Trade.* Cambridge: MIT Press.

Fujita M, Thisse J. 2002. *Economics of Agglomeration.* Cambridge: Cambridge University Press.

Fujita M, Thisse J. 2009. New economic geography: An appraisal on the occasion of Paul Krugman's 2008 Nobel Prize in Economic Sciences. *Regional Science and Urban Economics* **39**: 109–199.

Garcia A. 2006. Estimating krugmans enomomic geography model for the spanish regions. *Spanish Economic Review* **8**: 83–112.

Graves P. 1976. A reexamination of migration, economic opportunity and the quality of life. *Journal of Regional Science* **16**: 107–112.

Graves P. 1979. A life-cycle empirical analysis of migration and climate by race. *Journal of Urban Economics* **6**: 135–147.

Graves P. 1980. Migration and climate. *Journal of Regional Science* **20**: 227–237.

Graves P, Arthur M, Sexton R. 1999. Amenities and the labor earnings funciton. *Journal of Labor Research* : 367–379.

Hanson G. 2005. Market potential, increasing returns and geographic. *Journal of International Economics* **67**: 1–24.

Head K, Mayer T. 2000. Non-Europe: the magnitude and causes of market fragmentation in Europe. *Weltwirtschaftliches Archiv* **136**: 285–314.

Helpman E. 1998. The size of regions. In Pines D, Zadka E, Zilcha I (eds.) *Topics in Public Economics*. Cambridge: Cambridge University Press, 33–54.

Henderson V. 2003. The urbanization process and economic growth: the so-what question. *Journal of Economic Growth* **8**: 47–71.

Hering L, Poncet S. 2010. Market access and individual wages: evidence from China. *The Review of Economics and Statistics* **92**: 145–159.

Iturra V, Paredes D. 2011. Construction of a spatial housing price index by estimating an almost ideal demand system.
URL `"http://ideas.repec.org/p/pra/mprapa/30581.html"`

Kanbur R, Venables A. 2005. patial inequality and development overview of unu-wider project.
URL `http://www.arts.cornell.edu/poverty/kanbur/WIDERProjectOverview.pdf`

Kiso T. 2005. Does new economic geography explain the spatial distribution of wages in japan. *mimeo* .

Krugman P. 1991. Increasing returns and economic geography. *Journal of Political Economy* : 483–499.

Krugman P. 1993. First nature, second nature, and metropolitan location. *Journal of Regional Science* **33**: 129–144.

Marshall A. 1890. *Principles of Economics*. New York: Prometheous Books.

Mills E, Simenauer R. 1996. New hedonic estimates of regional constant quality house prices. *Journal of Urban Economics* **39**: 209–215.

Mion G. 2004. Spatial externalities and empirical analysis: the case of Italy. *Annals of the Association of American Geographers* **56**: 97–118.

Morande F, Soto R, Vergara R. 1997. Achilles, the tortoise and regional growth in chile. In Morande F, Vergara R (eds.) *Estudios Empricos sobre el Crecimiento en Chile*. 141–170.

Myrdal G. 1957. *Economic Theory and Underveloped Regions*. Methuen: University Paperbacks.

Paredes D, Aroca P. 2008. Metodologa para estimar un indice regional de costo de vivienda en Chile. *Latin American Journal of Economics* **45**: 129–143.

Partridge M. 2010. The duelling models: Neg vs amenity migration in explaining us engines of growth. *Papers in Regional Science* **89**: 513–536.

Partridge M, Rickman D, Ali K, Olfert M. 2009. Agglomeration spillovers and wage and housing cost gradients across the urban hierarchy. *Journal of International Economics* **78**: 126–140.

Puga D. 1998. Urbanisation patterns: European versus less developed countries. *Journal of Regional Science* **38**: 231–252.

Puga D. 1999. The rise and fall of regional inequalities. *European Economic Review* **43**: 303–334.

Puga D. 2002. European regional policies in the light of recent location theories. *Journal of Economic Geography* **2**: 372–406.

Redding S. 2010. The empirics of new economic geography. *Journal of Regional Science* **50**: 297–311.

Roback J. 1982. Wages, rents, and the quality of life. *Journal of Polical Economy* **90**: 1257–1278.

Roback J. 1988. Wages, rents, and amenities: Differences among workers and regions. *Economic Inquiry* **26**: 23–41.

Rosembaum P, Rubin D. 1985. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician* **39**: 33–38.

Rosen R. 1974. Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy* **82**: 34–55.

Rosenbaum P, Rubin D. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**: 41–55.

Rosenthal S, Strange W. 2004. Evidence on the nature and sources of agglomeration economies. In Henderson J, Thisse J (eds.) *Handbook of Regional and Urban Economics*. 2120–2171.

Rubin D. 1976. Matchingmethods that are equal percent bias reducing: some examples. *Biometrics* **32**: 109–120.

Rubin D, Thomas N. 2000. Combining propensity score matching with additional adjustments for prognostic covariates. *Journal of American Statistical Association* **95**: 571–585.

Samuelson P. 1952. The transfer problem and transport costs: The terms of trade when impediments are absents. *Economic Journal* : 278–304.

Smith H. 1997. matching with multiple controls to estimate treatment effects in observational studies. *Sociological Methodology* **27**: 325–353.

Sudekum J. 2006. Agglomeration and regional costs of living. *Journal of Regional Science* **46**: 529–543.

Thibodeau T. 1989. Housing price indexes from the 1974-1983 simsa annual housing surveys. *Real Estate Economics* **1**: 100–117.

Thisse J. 2010. Toward a unified theory of economic geography and urban economics. *Journal of Regional Science* **50**: 281–296.

Tobler W. 1970. A computer movie simulating urban growth in the Detroit region. *Economic Geography* : 234–240.

Tritchler D. 1995. Interpreting the standardized difference. *Biometrics* **51**: 351–353.

Venables A. 2005. Spatial disparities in developing countries: cities, regions and international trade. *International Economic Review* **5**: 3–21.